

Tilburg University

Kwaliteit van de juridische annotatie

van Dijck, G.

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
van Dijck, G. (2011). *Kwaliteit van de juridische annotatie: Een empirische studie naar kenmerken en kwaliteitsindicatoren*. (Juridische Methodologie). Boom Juridische Uitgevers.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

De kwaliteit van de juridische annotatie

Universiteit van Tilburg
Research Group for Methodology of Law and Legal Research

De kwaliteit van de juridische annotatie

*Een empirische studie naar kenmerken
en kwaliteitsindicatoren*

Gijs van Dijck

Boom Juridische uitgevers
Den Haag
2011

Omslagontwerp: Haagsblauw, Den Haag

© 2011 G. van Dijck / Boom Juridische uitgevers

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3060, 2130 KB Hoofddorp, www.reprorecht.nl). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (art. 16 Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 978-90-8974-575-0

NUR 820

www.bju.nl

Inhoudsopgave

Voorwoord	1
Hoofdstuk 1	3
Inleiding en verantwoording	3
1 Introductie	3
2 Achtergrond	3
3 Eerdere studies naar annotaties	6
4 Onderzoeksvraag	8
5 Werkwijze	9
Hoofdstuk 2	11
Deskundigenoordeel	11
1 Inleiding	11
2 Overeenstemming kwaliteit van annotaties	12
3 Inventarisatie mogelijke kwaliteitsindicatoren	16
4 Relatie mogelijke kwaliteitsindicatoren – beoordelingen van annotaties	19
5 Conclusie	24
Hoofdstuk 3	26
Kwaliteitsindicatoren blijkend uit annotaties	26
1 Inleiding en methode	26
2 Kenmerken van annotaties	28
3 Observaties en voorlopige conclusies	31
3.1 Kwaliteitsindicatoren	32
3.2 Kenmerken met geen of zeer geringe invloed	33
3.3 Kenmerken met ambigu beeld	35
4 Resultaten fuzzy-set qualitative comparative analysis (fsQCA)	35
5 Conclusie	38

Hoofdstuk 4 _____ 41

Experimentele studies naar de kwaliteit van annotaties _____ 41

1	Inleiding	41
2	Studie 1 – Schetsen van implicaties en de kwaliteit van annotaties	41
2.1	Onderzoeksvraag en opzet	41
2.2	Methode	42
2.3	Resultaten	45
2.4	Discussie	51
3	Studie 2 – Toetsen aan ‘hoger’ recht en de kwaliteit van annotaties	52
3.1	Onderzoeksvraag en opzet	52
3.2	Methode	52
3.3	Resultaten	54
3.4	Discussie	57
4	Conclusie	58

Hoofdstuk 5 _____ 59

Wetenschappelijkheid van annotaties _____ 59

1	Inleiding	59
2	Wetenschappelijkheid van juridisch onderzoek	60
3	Kritisch rationalisme	63
4	Kuhns paradigmatische theorie	66
5	Lakatos	70
6	Multicriteria-benadering	71
7	Tussenstand: geen criteria, wel gemeenschappelijke kenmerken	74
8	Repliceerbaarheid/herhaalbaarheid	74
8.1	Beschrijven of verklaren van regelmatigheden of patronen	75
8.2	Evaluatieve (normatieve) annotaties	79
9	Andere kenmerken van pseudowetenschappelijkheid en hun relatie tot annotaties	83
10	Experimentele studie naar effect van aanleggen wetenschappelijke criteria op beoordeling van annotaties	86
10.1	Onderzoeksvraag en opzet	86
10.2	Methode	86
10.3	Resultaten	88
10.4	Discussie	90
11	Conclusie	91

Hoofdstuk 6 _____ 94

Naar een beoordeling van de (academische) kwaliteit _____ 94

1	Inleiding	94
2	Procedurele waarborgen	94
3	Inhoudelijk criterium: normatieve diepgang	95
4	Criteria omtrent de werkwijze	97
5	Status van annotaties	100

Overzicht van verkort aangehaalde literatuur	104
Bijlage 1	111
Betrouwbaarheid vragenlijst & selectie annotaties	111
1 Meten van betrouwbaarheid	111
2 Selectie van annotaties en beoordelaars	111
3 Uitkomst betrouwbaarheidsonderzoek	112
Bijlage 2	114
Overzicht scores potentiële kwaliteitsindicatoren	114
Bijlage 3	120
Globale samenvattingen van de gecodeerde annotaties	120
1 Bestuursrecht	120
2 Privaatrecht	124
3 Strafrecht	130
Bijlage 4	135
Systematische analyse van annotaties (per rechtsgebied)	135
1 Analyse van bestuursrechtelijke annotaties	135
2 Analyse van privaatrechtelijke annotaties	146
3 Analyse van strafrechtelijke annotaties	158

VOORWOORD

De kwaliteit en status van annotaties, en juridisch onderzoek in het algemeen, heeft de laatste jaren sterk in de belangstelling gestaan. In dit verband zijn verschillende opvattingen en ideeën naar voren gebracht over wat de kenmerken van annotaties zijn, en of zij als academisch product en wetenschappelijk zouden moeten worden gekwalificeerd, vooral met het oog op onderzoeksvisitaties.

Deze studie betreft een ‘evidence based’ onderzoek naar kwaliteitsindicatoren voor annotaties. Er zijn verschillende empirische methoden gebruikt, te weten het bevragen van deskundigen, het analyseren van een hoeveelheid annotaties en het uitvoeren van experimenten. Dit onderzoek reikt specifiekere en soms andere criteria aan dan die in de literatuur zijn genoemd. Een verkenning van wat onderzoek tot wetenschappelijk onderzoek maakt, geeft zicht op de wetenschappelijkheid van annotaties en de status die zij zouden moeten krijgen.

De resultaten zijn van belang voor annotatoren die op hun annotaties willen reflecteren en voor degenen die annotaties willen leren schrijven of daarover doceren. Daarnaast zijn de resultaten relevant voor het beoordelen van annotaties, bijvoorbeeld in onderzoeksvisitaties, alsook voor de status van annotaties en voor de kenmerken van juridisch onderzoek in het algemeen.

Ik dank de zeven deskundigen die de annotaties hebben geselecteerd en beoordeeld voor hun medewerking aan het onderzoek. Datzelfde geldt voor de deelnemers aan de experimentele studies. Ik noem ze niet bij naam in verband met de anonimiteit die is beloofd. Heinze Oost hielp bij het opzetten van het empirische deel van de studie. Hij mag het resultaat helaas niet meemaken. Rob van Gestel, Reinout Wibier en Jan Vranken hebben een concept van deze studie van commentaar voorzien. Tot slot is ook aan Paul Marcelis een woord van dank verschuldigd. Hij leverde de noodzakelijke ondersteunende werkzaamheden.

Tilburg Law School en de *Research Group for Methodology of Law and Legal Research* maakten het financieel mogelijk deze studie uit te voeren.

Gijs van Dijk
Oktober 2011

Inleiding en verantwoording

1 Introductie

Een annotatie bestaat doorgaans uit een bespreking van een uitspraak van een rechter. Hoewel de annotatie is ingeburgerd in het juridisch forum in Nederland, staat deze publicatievorm echter onder druk. Bij onderzoeksvisities in Nederland worden annotaties in beginsel niet als academisch product aangemerkt, terwijl dat voorheen wel het geval was. Meer in het algemeen zijn twijfels geuit over de aard, het nut en de wetenschappelijkheid van het bestaande juridisch onderzoek.

Dit onderzoek analyseert wat een annotatie tot een goede annotatie maakt. Gekozen is voor een ‘bottom-up approach’, waarbij de kwaliteitsindicatoren zijn achterhaald door deskundigen te bevragen, door goede annotaties met minder goede annotaties te vergelijken en door experimenten te verrichten. Het onderzoek geeft zicht op de mate van overeenstemming tussen beoordelaars over de kwaliteit van annotaties en de te hanteren beoordelingsmaatstaven. Het achterhalen van kwaliteitsindicatoren geeft inzicht in de aard en het karakter van annotaties, en daarmee mogelijk ook in de aard van juridisch-dogmatisch onderzoek. Voorts zijn deze indicatoren van belang voor de wijze waarop annotaties worden beoordeeld en voor de kwalificatie die aan deze beoordeling wordt verbonden.

Het achterhalen van kenmerken van goede/uitstekende annotaties gebeurt aan de hand van empirisch onderzoek, meer specifiek vragenlijsten, document-analyse en experimentele studies. Daarmee is deze studie een van de weinige studies die de kwaliteit en kenmerken van annotaties op empirische wijze in kaart brengt.

2 Achtergrond

Voor Nederland kan de opkomst van de annotatie worden geplaatst rond het jaar 1850.¹ Het annoteren van rechterlijke uitspraken vindt vermoedelijk zijn oorsprong in Frankrijk, in het midden van de negentiende eeuw, toen tijdschriften werden opgericht die speciaal waren bedoeld om gerechtelijke uitspraken te signaleren en te becommentariëren.² Annotaties genieten in Frankrijk een hoog

¹ Jansen 2003. Zie voorts Bleeker 2010, die de eerste bouwrechtelijke noot aantreft in het jaar 1925.

² Duxbury 2001, p. 49, 51, onder verwijzing naar een Franstalige bron.

aanzien.³ Zij worden beschouwd als onderdeel van een feedbackproces voor gerechtelijke uitspraken.⁴ De nauwe band met de rechtspraak maakt dat annotaties in belangrijke mate invloed hebben op de ontwikkeling van het recht,⁵ hoewel de relatie minder groot is dan soms wel wordt gedacht, gegeven de hoeveelheid zaken die rechters hebben, het gebrek aan tijd om publicaties goed bij te houden en omdat annotatoren soms moeten gissen naar de achterliggende gedachten bij bepaalde uitspraken.⁶

Inmiddels staan annotaties onder druk. Een aantal, met elkaar samenhangende ontwikkelingen illustreert dit, waarbij oorzaak en gevolg niet altijd goed zijn te onderscheiden. Een eerste ontwikkeling is de explosieve toename van het aantal tijdschriften met een annotatierubriek. Behalve in tijdschriften als *Nederlandse Jurisprudentie (NJ)* en *Administratiefrechtelijke Beslissingen (AB)*, verschijnen annotaties thans ook in tal van andere tijdschriften, waaronder *TvI*, *OR*, *JA*, *JAR*, *JOR*, *NbBW*, *NTBR*, *JB*, aan te vullen met een hoeveelheid internationale en andere Nederlandstalige tijdschriften die annotaties publiceren. Door de opkomst van tijdschriften met annotaties rijzen vragen ten aanzien van de selectie en kwaliteit ervan. In sommige tijdschriften beperkt een annotatie zich tot een samenvatting. In andere annotaties worden uitspraken geëvalueerd of worden vergelijkingen getrokken met andere rechtsgebieden of met andere rechtstelsels.

Een tweede ontwikkeling raakt de nationale en casusgebonden oriëntatie van annotaties. Deze staat op gespannen voet met de ontwikkeling dat van juridisch onderzoek in toenemende mate wordt verwacht dat het internationaal georiënteerd is of een interdisciplinair dan wel een multidisciplinair karakter heeft. Voorts – de derde ontwikkeling – concurreert juridisch onderzoek voor de verdeling van onderzoeksgelden (NWO) met wetenschapsdisciplines die meestal een sociaalwetenschappelijke achtergrond hebben. Deze disciplines, alsook de beoordelaars, (er)kennen de annotatie niet als publicatievorm.

Hiermee hangt samen het verminderde aanzien van de annotatie. Vranken brengt deze ontwikkeling in verband met de dienstbaarheid van de rechtswetenschap aan de praktijk, wat hij beschouwt als een belangrijk kenmerk van de rechtswetenschap.⁷ De praktische oriëntatie van annotaties komt volgens Vranken tot uitdrukking in vijf elementen, te weten (1) een analyse en uitleg van wat is beslist, (2) het blootleggen van achtergronden, (3) het zoeken naar verbanden met eerdere uitspraken of andere delen van het recht dan wel het beoordelen van

³ Duxbury 2001, p. 54.

⁴ Bell & Boyron & Whittaker 1998, p. 34; Duxbury 2001, p. 53.

⁵ Duxbury 2001, p. 52, met voorbeelden.

⁶ Tunc 1976, p. 472.

⁷ Vranken 2005. De auteur stelt deze dienstbaarheid ter discussie en pleit kort gezegd voor juridisch onderzoek dat minder op de rechtspraak is georiënteerd.

de uitspraak in het licht van de relevante leerstukken en deze, als de uitspraak daar aanleiding toe gaf, aanvullen of verfijnen, (4) het aandragen van alternatieve oplossingen en (5) het vooruitdenken hoe toekomstige, vergelijkbare of aanpalende gevallen beslist moeten worden.⁸ Omdat ook de dienstbaarheid aan de praktijk onder druk staat, verrast het volgens hem niet dat ook de typische publicatie van de rechtswetenschap – de annotatie – onder druk staat.

De vijfde ontwikkeling betreft de beslissing in eigen kring om annotaties niet langer te kwalificeren als academische publicaties. Voor het eerst gebeurde dit in de onderzoeksvisitatie die in 2002 is gehouden.⁹ In het verlengde hiervan ligt het advies van de *Commissie Voorbereiding Onderzoeksbeoordeling Rechtsgeleerdheid*, welke commissie als opdracht had een disciplineprotocol te ontwerpen voor de rechtswetenschap en criteria te formuleren voor de beoordeling van juridisch onderzoek. In haar rapport *Oordelen over rechten* (2005) adviseert zij de annotatie niet als academische publicatievorm aan te merken, tenzij de auteur weet uit te leggen dat de publicaties als academisch dienen te worden aangemerkt. Aan deze opvatting ligt de gedachte ten grondslag dat annotaties divers van aard zijn en lang niet altijd zullen bijdragen aan ‘increasing of the body of academic knowledge’, terwijl waar wél sprake is van originaliteit het vaak gaat om aanzetten tot een bijdrage aan de theorievorming, aanzetten die dan ook vaak door de annotator in opvolgende, uitvoeriger (wetenschappelijke) publicaties worden uitgewerkt.¹⁰ Tot slot, zo stelt het rapport, kennen vrijwel alle juridische tijdschriften annotatierubrieken, waarmee een voldoende wetenschappelijk gehalte niet zonder meer een gegeven is, en is een beperking naar tijdschrift vooralsnog lastig te motiveren.

Een laatste, zesde ontwikkeling waardoor annotaties vermoedelijk onder druk zijn komen te staan, houdt verband met de discussie over de wetenschappelijkheid van juridische publicaties in het algemeen. Kritiek op juridische publicaties raken annotaties in het bijzonder, nu deze kunnen worden gezien als een publicatievorm die illustratief is voor de wijze waarop juridisch onderzoek wordt gedaan.¹¹

Het meest kritisch (in Nederland) over de wetenschappelijkheid van juridisch onderzoek lijkt De Geest te zijn geweest.¹² Hij vergelijkt juridisch onderzoek met de natuurwetenschappelijke onderzoekscyclus, wijst op methodologische fouten in juridisch onderzoek, en komt uiteindelijk tot de conclusie dat juridisch onderzoek te veel scholastisch is en te weinig wetenschappelijk. Ande-

⁸ Vranken 2005, nr. 141.

⁹ Ook daarvoor stond de annotatie onder druk, waarover onder meer Barendrecht 1996, met verdere verwijzingen.

¹⁰ Zo ook Hondius 2010.

¹¹ Vranken 2005, p. 144.

¹² De Geest 2004.

ren zijn optimistischer. Zij menen dat het bestuderen en becommentariëren van wat de rechter doet, te kwalificeren is als wetenschapsbeoefening.¹³ Niet zonder meer, want de onderzoeker moet expliciteren dat hij dogmatisch onderzoek doet en geen metajuridisch perspectief kiest of gebruikt. Voorts is opgemerkt dat juridisch onderzoek zou moeten werken aan zijn toetsbaarheid (openheid, controleerbaarheid en falsificeerbaarheid), aan vernieuwing, vooruitgang, durf en ambitie.¹⁴ Zo zou onderzoek minder op de praktijk moeten zijn georiënteerd en zou er meer aandacht moeten zijn voor de ‘waaromvraag’ in onderzoek, wat wil zeggen dat de onderzoeker dient aan te geven wat de waarde is van het onderzoek dat hij heeft uitgevoerd, daarbij inbegrepen de vraag welk werk anderen al hebben verricht.¹⁵

De beslissing om annotaties niet langer als wetenschappelijk product aan te merken, in combinatie met de discussie over de wetenschappelijkheid van juridisch onderzoek in het algemeen, lijkt voor een aantal vooraanstaande Nederlandse juristen aanleiding te zijn geweest om de wetenschappelijkheid van de annotatie te bepleiten. De argumenten die in dit verband zijn aangevoerd, lopen uiteen. Sommigen zoeken de wetenschappelijkheid in de omstandigheid dat juristen met het analyseren van praktijkgevallen bijdragen aan het inzicht in leerstukken en daarmee een wetenschappelijk doel nastreven, te weten kennisvergroting.¹⁶ Een andere opvatting is dat de wetenschappelijkheid schuilt in de kenmerken van de annotatie: het uitleggen, analyseren en plaatsen van de uitspraak en haar gevolgen.¹⁷ Verder wordt gesteld dat annotaties kunnen voldoen aan eisen als onafhankelijkheid, openheid, eerlijkheid, helderheid, eenvoud en nauwgezetheid, en daarmee wetenschappelijk zijn.¹⁸ Dit zou aansluiten bij wat annotatoren, historisch gezien, altijd al hebben gedaan. Juristen zouden met het schrijven van annotaties oorspronkelijk niet alleen uit zijn op een beschrijving van het bestaande recht, maar ook kennisvergroting hebben nagestreefd: het analyseren van praktijkgevallen zou bijdragen aan het ontwikkelen van het inzicht in (privaatrechtelijke) leerstukken.¹⁹

3 Eerdere studies naar annotaties

Empirisch onderzoek dat in het verleden naar annotaties is gedaan, beperkt zich, voor zover ik heb kunnen overzien, tot een studie van Crombag en Cohen. Zij

¹³ Zie bijvoorbeeld Van Rhee 2004.

¹⁴ Stolker 2003, p. 774-776.

¹⁵ Vranken 2005, p. 150-154.

¹⁶ Jansen 2003.

¹⁷ Sniijders 2003.

¹⁸ Bloembergen 1988; Bloembergen 2002.

¹⁹ Jansen 2003.

hebben, na een empirische (voor)studie, hypotheses opgesteld over wat de inhoud van een annotatie bepaalt.²⁰ Daartoe hebben zij de werkwijze van A.R. Bloembergen uitgebreid beschreven, die destijds annotator was. Zij vroegen hem zijn denkwerk en het analyseproces bij het schrijven van een annotatie zo precies mogelijk in te spreken op een taperecorder. Het idee hierachter was dat een dergelijke beschrijving didactische relevantie zou hebben en een bijdrage zou leveren aan de studie van juridische methodologie: gaan onder de uniekheid van annotaties regelmatigheid schuil?

Ondanks de beperkte steekproef – één annotator – leverde het beschrijven van ongeveer 36 uur ‘annoteren’ ten minste een drietal belangrijke inzichten op. Deze inzichten benadrukken de factor ‘toeval’ bij de inhoud van de annotatie. Zo berust, volgens het onderzoek, de keuze van literatuur ten dele op invallen, waarbij het gaat om in het nabije verleden gepubliceerde literatuur of om literatuur waarmee de annotator vertrouwd is geraakt.

Ook leverde het onderzoek twee hypotheses op. De eerste is dat de ideeën en invalshoeken van noten sterk afhankelijk zijn van de mate waarin en de breedte waarmee de annotator de literatuur bijhoudt. De tweede hypothese is dat verschillen tussen annotatoren voor een belangrijk deel zijn terug te voeren op verschillen in de verzameling bronnen die zij paraat en tot hun beschikking hebben.

Het was de bedoeling dat meer protocollen van dezelfde annotator en protocollen van andere annotatoren over hetzelfde arrest zouden worden verzameld. Daarmee zou inzicht kunnen worden verkregen in wat de constante factoren zijn, dat wil zeggen factoren die losstaan van de persoon van de annotator. Zo ver is het echter niet gekomen. Navraag leert dat dit niet zozeer te maken had met de tijd die het onderzoek in beslag nam – de verwachting was dat het verzamelen en analyseren van acht tot tien protocollen een onderzoeker gedurende meer dan een jaar een goedgevulde dagtaak zou geven – maar veel meer met het gebrek aan bereidwilligheid van de toenmalige NJ-annotatoren om mee te werken aan het onderzoek. Inmiddels – het onderzoek stamt uit 1974 – lijken sommige hypotheses gedateerd, met name vanwege de opkomst van informatietechnologie en internet. In vergelijking met 1974 zijn juridische bronnen, ook buitenlandse, thans sneller en eenvoudiger te achterhalen.

In de internationale literatuur heeft de annotatie als publicatievorm in met name Frankrijk aandacht gekregen. Daar valt op de uitstekende reputatie die annotaties genieten.²¹ Deze wordt deels toegeschreven aan de historisch hiërarchische structuur binnen het Franse forum van juristen.²² Minstens zo belangrijk echter is de korthed waarmee Franse rechters, met name de *Cour de Cassation*,

²⁰ Crombag & Cohen 1974.

²¹ Duxbury 2001, p. 54, p. 57.

²² Duxbury 2001, p. 57-58.

hun beslissingen motiveren, wat voor annotatoren de mogelijkheid geeft om de relatie met bestaande jurisprudentie te onderzoeken en regelmatigheden af te leiden uit een hoeveelheid annotaties.²³

Het contrast met bijvoorbeeld de Verenigde Staten en Engeland is groot. In de Verenigde Staten genieten annotaties over het algemeen weinig aanzien en worden zij vooral geschreven door studenten.²⁴ Daar beperkt het belang van annotaties zich tot het onderwijs, waarin studenten in het kader van hun opleiding soms annotaties dienen te schrijven. De reputatie van annotaties in Engeland is niet veel beter.²⁵

Uit het voorgaande volgt dat de waardering van annotaties deels historisch is gegroeid, maar ook in belangrijke mate afhankelijk is van institutionele factoren en dan met name de wijze waarop en de mate waarin rechters hun beslissingen motiveren.

4 Onderzoeksvraag

Dit onderzoek richt zich op de kwaliteit van annotaties. Dit kan op twee niveaus. Het eerste niveau is dat op het niveau van verschillen in waardering tussen landen of systemen, zoals hiervoor beschreven. Dit onderzoek concentreert zich op het andere niveau, te weten het niveau van verschillen tussen individuele annotaties. Wat in dit verband opvalt, is dat in de discussie over de kwaliteit en status van annotaties algemene uitspraken worden gedaan over annotaties, met name over de kenmerken en wetenschappelijkheid ervan. Deze aanpak miskent de mogelijke kwaliteitsverschillen tussen annotaties en typen annotaties, en mist bovendien feitelijke onderbouwing. Dit onderzoek richt zich op het scherp krijgen van verschillen tussen individuele annotaties en op het aan de hand daarvan achterhalen van criteria op basis waarvan verschillen in kwaliteit kunnen worden achterhaald:

Wat zijn kwaliteitsindicatoren van annotaties?

Beantwoording van deze vraag draagt bij aan het verkrijgen van een beter inzicht in hoe kwalitatief goede of uitstekende annotaties kunnen worden onderscheiden van matige of slechte annotaties. Deze informatie draagt tevens bij aan het achterhalen van de kenmerken van juridisch onderzoek, in het bijzonder juridisch-dogmatisch onderzoek. Daarnaast geeft dit onderzoek meer zicht op het wetenschappelijke karakter van annotaties, en mogelijk ook op het karakter van juridisch-dogmatisch onderzoek in het algemeen. De annotatie is immers

²³ Goutal 1976, p. 54-65; Duxbury 2001, p. 53.

²⁴ Duxbury 2001, p. 56.

²⁵ Duxbury 2001, p. 55-56.

een voorbeeld van juridisch-dogmatisch onderzoek. Onderzoek naar de kwaliteit en de methodologie ervan kan daarmee tevens relevant zijn voor de methodologie en kwaliteit van juridisch-dogmatisch onderzoek in het algemeen.

In de discussie over de aard, het nut en de status van annotaties (en juridisch onderzoek in het algemeen) zijn annotaties in verband gebracht met eisen van wetenschappelijkheid. Deze studie staat daarom tevens stil bij de betekenis van de resultaten voor de (on)wetenschappelijkheid en status van annotaties.

5 Werkwijze

Verschillende gegevens en analysetechnieken worden gebruikt om criteria op te sporen die verklaren wat een annotatie tot een goede annotatie maakt. De eerste stap is, kort gezegd, om aan deskundigen te vragen wat een annotatie tot een goede annotatie maakt (Hoofdstuk 2). Dit gebeurt aan de hand van vragenlijsten. Deze analyse wordt gevolgd door een analyse van annotaties, waarbij goede met minder goede annotaties worden vergeleken en wordt gezien waardoor verschillen in kwaliteit kunnen worden verklaard (Hoofdstuk 3). Deze analyses zullen het echter niet mogelijk maken om ‘harde’, causale uitspraken te doen over (directe) oorzaken en gevolgen. Om die reden wordt een aantal experimenten uitgevoerd (Hoofdstuk 4). Een uitgebreidere verantwoording van de analyses die zijn uitgevoerd vindt steeds plaats aan het begin van het betreffende hoofdstuk.

Na criteria te hebben geïdentificeerd, wordt ingegaan op de wetenschappelijke status van annotaties. Meer in het bijzonder wordt stilgestaan bij de bijdrage van de onderzoeksresultaten aan het bepalen van het wetenschappelijke karakter van annotaties (Hoofdstuk 5). Het identificeren van de eisen om te bepalen of onderzoek, juridisch onderzoek in het bijzonder, wetenschappelijk is, gebeurt aan de hand van een literatuurstudie. Afgesloten wordt met een verkenning van de gevolgen van deze studie voor de beoordeling van annotaties (Hoofdstuk 6).

Deskundigenoordeel

1 Inleiding

Om kwaliteitsverschillen tussen annotaties te achterhalen, is gekozen voor een opzet waarin aan deskundigen is gevraagd om goede/uitstekende van matige/slechte annotaties te onderscheiden en hun beoordelingen te motiveren. Daartoe is een hoeveelheid annotaties geselecteerd voor een drietal rechtsgebieden waarin geregeld annotaties worden geschreven, te weten het bestuursrecht, het privaatrecht en het strafrecht.

De annotaties zijn geselecteerd door hoogleraren die deskundig zijn op het terrein van respectievelijk het bestuursrecht, privaatrecht en strafrecht. Het verzoek was om zestien annotaties te selecteren, waarvan er acht van matige/slechte kwaliteit waren en acht van goede/uitstekende kwaliteit. Dit betekent dat van de geselecteerde annotaties ongeveer de helft van slechte/matige kwaliteit was.²⁶ De overige annotaties waren als goed/uitstekend beoordeeld. Het totaal aantal annotaties dat is aangeleverd, is 50, waarvan 18 bestuursrechtelijke, 16 privaatrechtelijke en 16 strafrechtelijke annotaties.

De annotaties zijn vervolgens geanalyseerd door één of twee andere deskundigen (hoogleraar of universitair hoofddocent) per rechtsgebied. Dit gebeurde aan de hand van een vooraf opgestelde vragenlijst, waarover meer in het vervolg van dit hoofdstuk. Zij beoordeelden de annotaties onafhankelijk van elkaar en waren ten tijde van het invullen van de vragenlijst niet bekend met de waardering zoals die in de selectie was gemaakt. Aan de hand van de ingevulde vragenlijsten kon worden bekeken in hoeverre de kwaliteitsoordelen (beoordelingen) overeenstemmen en konden de redenen worden achterhaald die aan de beoordelingen ten grondslag lagen.

De opzet in dit hoofdstuk is als volgt. Eerst wordt een indruk verkregen van de mate waarin er overeenstemming is tussen de beoordelaars over de kwaliteit van annotaties (par. 2). Daarna wordt getoetst welke criteria beoordelaars aanleggen om te bepalen of annotaties moeten worden gekwalificeerd als goed/uitstekend of als slecht/matig, en of er overeenstemming is over de criteria om annotaties te beoordelen (par. 3-4).

²⁶ Een deelnemer aan het onderzoek spreekt in dit verband van ‘minder geslaagde annotaties’.

2 Overeenstemming kwaliteit van annotaties

De eerste stap was om voor de verschillende rechtsgebieden (bestuursrecht, privaatrecht, strafrecht) na te gaan in hoeverre er overeenstemming bestond tussen de beoordelaars over de kwaliteit van de annotaties. De kwaliteitsoordelen, gemeten aan de hand van een vierpuntschaal (slecht/matig/goed/uitstekend) zijn omgezet naar dichotome variabelen (slecht/matig versus goed/uitstekend). Hierdoor kon worden berekend in hoeverre er overeenstemming bestond tussen de beoordelaars over de kwaliteit van de annotaties. De overeenstemming is zowel berekend in termen van percentages als aan de hand van de zogenoemde kappa (κ), een maat om de samenhang te berekenen.

Indien zowel het hokje matig als goed was aangevinkt, zijn deze annotaties niet in de berekening meegenomen. Verder zijn drie bestuursrechtelijke annotaties en één privaatrechtelijke annotatie niet meegenomen in de analyse. Reden hiervoor is dat twee (bestuursrechtelijke) annotaties door één van de beoordelaars bleek te zijn geschreven, wat de beoordeling ervan mogelijk onbetrouwbaar maakt. De andere bestuursrechtelijke annotatie betreft geen commentaar bij een rechterlijke uitspraak, maar bij een onderzoeksrapport. Voor de privaatrechtelijke annotatie die niet is meegenomen geldt dat twee van de drie beoordelaars twijfels hebben geuit ten aanzien van het antwoord op de vraag of de annotatie wel een annotatie is (en niet een artikel).

De resultaten laten zien dat de mate van overeenstemming tussen de annotaties die als slecht/matig zijn gekwalificeerd en de annotaties die als goed/uitstekend zijn aangemerkt, varieert tussen de 50% en 86%. Ten aanzien van de bestuursrechtelijke annotaties is een mate van overeenstemming van 60% gevonden (Tabel 1). Deze mate van overeenstemming is vanuit statistisch oogpunt niet als betrouwbaar aan te merken ($\kappa = .20$, $p < .45$).

Tabel 1 – Mate van overeenstemming kwaliteit bestuursrechtelijke annotaties

Categorie	Overeenstemming B1 - B2
Slechte/matige annotaties	4 annotaties
Goede/uitstekende annotaties	5 annotaties
Totaal	60% (9 van de 15)
K	.20

Noot: B1 = Beoordelaar 1. B2 = Beoordelaar 2.

Box 1 – Kappa

De kappa (κ) is een statistische maat waarmee de mate van overeenstemming tussen beoordelaars kan worden gemeten. Als vuistregel geldt dat een kappa-waarde tussen de .40 en .59 als matige overeenstemming wordt gezien, tussen de .60 en .79 als substantieel, en waarden van .80 of hoger als uitstekend (zie **Landis & Koch 1977**, p. 165). Daarbij komt dat de samenhang statistisch significant moet zijn ($p < .05$), wat inhoudt dat de kans klein is dat de gevonden samenhang op toeval berust.

De overeenstemming tussen de beoordelaars ten aanzien van de strafrechtelijke annotaties is groter ($\kappa = .50$, $p = .02$), hoewel deze overeenstemming statistisch gezien matig is (Tabel 2).

Tabel 2 – Mate van overeenstemming kwaliteit strafrechtelijke annotaties

Categorie	Overeenstemming B1 - B2
Slechte/matige annotaties	4 annotaties
Goede/uitstekende annotaties	8 annotaties
Totaal	75% (12 van de 16)
K	.50*

Noot: * $p < .05$. B1 = Beoordelaar 1. B2 = Beoordelaar 2.

Voor de privaatrechtelijke annotaties is de mate van overeenstemming twee keer berekend. Dit kon, omdat de beoordelaars in het begin van het onderzoek zijn betrokken bij het opstellen en verfijnen van de vragenlijsten. In dat verband zijn zij ook in het beginstadium voor sommige (maar niet voor alle annotaties) gevraagd om een kwaliteitsoordeel uit te spreken over de annotaties die zij hebben gelezen.²⁷

Tussen de eerste en de tweede beoordeling zat een periode van een aantal maanden. Aan twee van de drie privaatrechtelijke beoordelaars is gevraagd om enkele maanden na het invullen van de vragenlijst deze vragenlijst nogmaals in te vullen en zich daarmee opnieuw uit te laten over de kwaliteit van de annotaties. Beide keren dienden zij op een vierpuntschaal aan te tekenen of zij de annotatie als slecht, matig, goed of uitstekend beoordeelden. De beoordelaars hadden bij de tweede beoordeling geen inzage in de vragenlijst die zij eerder hadden

²⁷ Er bestond geen reden om aan te nemen dat de beoordelaars tijdens een van de rondes de vragenlijsten minder serieus hebben ingevuld. Niet valt uit te sluiten dat het meerdere malen beoordelen van dezelfde annotatie andere effecten heeft, zoals het kritischer worden ten aanzien van de betreffende annotatie, zie ook hierna.

ingevuld, tenzij zij een kopie hadden gemaakt, maar daar zijn geen aanwijzingen voor.

De mate van overeenstemming en analyse van de mate van overeenstemming tussen de beoordelaars onderling geeft vergelijkbare percentages. Bij de eerste meting varieerden de percentages tussen 50% en 86%, waarbij met name de mate van overeenstemming tussen beoordelaar 1 en beoordelaar 2 hoog was, maar de mate van overeenstemming tussen beoordelaar 2 en 3 laag (Tabel 3).

Tabel 3 – Kwaliteit annotaties privaatrecht (eerste meting)

Categorie	Overeenstemming		
	B1 vs B2 (1 ^{ste} meting)	B1 vs B3 (1 ^{ste} meting)	B2 vs B3 (1 ^{ste} meting)
Slechte/matige annotaties	4 annotaties	2 annotaties	3 annotaties
Goede/uitstekende annotaties	8 annotaties	1 annotatie	2 annotaties
Totaal	86% (6 vd 7)	50% (3 vd 6)	71% (5 vd 7)
K	.70*	.00	.42

Noot: * $p < .05$. B1 = Beoordelaar 1. B2 = Beoordelaar 2. B3 = Beoordelaar 3.

Bij de tweede meting was er minder variatie. Daar lag de mate van overeenstemming tussen 62% en 75% (Tabel 4).

Tabel 4 – Kwaliteit annotaties privaatrecht (tweede meting)

Categorie	Overeenstemming		
	B1 vs B2 (2 ^e meting)	B1 vs B3 (2 ^e meting)	B2 vs B3 (2 ^e meting)
Slechte/matige annotaties	5 annotaties	6 annotaties	6 annotaties
Goede/uitstekende annotaties	3 annotaties	2 annotaties	3 annotaties
Totaal	75% (8 vd 12)	62% (8 vd 13)	64% (9 vd 14)
K	.33	.16	.29

Noot: * $p < .05$. B1 = Beoordelaar 1. B2 = Beoordelaar 2. B3 = Beoordelaar 3. De oordelen van B1 zijn dezelfde als in de eerste meting.

Wat opvalt, zijn de verschillen tussen de eerste en de tweede meting. Dit roept de vraag op hoe constant beoordelaars zijn als zij meerdere malen de kwaliteit van dezelfde annotaties beoordelen.

Aan de hand van de gegevens is onderzocht in hoeverre beoordelaar 2 en beoordelaar 3, want alleen zij waren bij de herhaalde meting betrokken, een

ander oordeel hadden bij de tweede meting in vergelijking met de eerste meting. Omdat in de eerste meting een beperkt aantal annotaties is beoordeeld, is ook het aantal annotaties in de tweede meting beperkt ($n=7$). Onderstaande kruistabel laat zien dat 29% van de beoordelingen in de tweede meting niet overeenkwam met de beoordeling in de eerste meting (71% overeenstemming) (Tabel 5). Verder laat de tabel zien dat er geen lijn te ontdekken is in de verschillen. Zo kwalificeert beoordelaar 1 een tweetal annotaties in eerste instantie als slecht/matig, maar daarna als goed/uitstekend, en geldt voor beoordelaar 3 het omgekeerde: een tweetal annotaties dat aanvankelijk als goed/uitstekend is aangemerkt, kwalificeert hij/zij in de volgende meting als slecht/matig.

Tabel 5 – Herhaalde meting kwaliteit annotaties (privaatrecht)

	Beoordelaar 2	Beoordelaar 3
Eerste keer slecht/matig, tweede keer slecht/matig	3 van de 7	4 van de 7
Eerste keer goed/uitstekend, tweede keer goed/uitstekend	2 van de 7	1 van de 7
Eerste keer slecht/matig, tweede keer goed/uitstekend	2 van de 7	0 van de 7
Eerste keer goed/uitstekend, tweede keer slecht/matig	0 van de 7	2 van de 7
Overeenstemming totaal	71% (5 uit 7)	71% (5 uit 7)
K	.33	.16
p-waarden	.15, <i>ns</i>	.21, <i>ns</i>

Noor: Voor de volledigheid zijn de kappa-waarden weergegeven. Gezien het lage aantal (< 9) en in het licht van voorgaande resultaten was het onwaarschijnlijk dat de toetsen significante resultaten zouden opleveren.

Hoewel intuïtief een hoge mate van overeenstemming zou worden verwacht tussen de eerste en de tweede meting, blijkt de herhaalde meting toch andere kwaliteitsoordelen op te leveren dan in eerste instantie het geval was.

De analyse geeft geen verklaring voor de afwijkingen. Voor de negatieve waarderings kan voortschrijdend inzicht (participatie aan het onderzoek) een verklaring zijn. Het bij herhaling lezen van dezelfde annotatie of het meerdere malen moeten doordenken van de annotatie in kwestie kan het kwaliteitsoordeel van die annotatie beïnvloeden. Ongeacht de verklaring suggereert de uitkomst dat niet alleen de subjectieve opvatting van de beoordelaar in kwestie over wat een annotatie tot een goede annotatie maakt van invloed is op het oordeel over een annotatie, maar ook de vraag of de beoordelaar de annotatie meerdere keren heeft bestudeerd.

Meer in het algemeen laten de resultaten zien hoe moeilijk het is om annotaties op een objectieve wijze te beoordelen. De omstandigheid dat beoordelaars, deskundig op hun rechtsgebied, zelf relatief laag scoren bij een herhaalde meting, geeft de moeilijkheid van een objectieve beoordeling al aan. Een beoordelaar merkte ook op dat het lastig is om uitspraken die op dezelfde materie betrekking hadden – uitleg van overeenkomsten in dit geval – goed te beoordelen. Volgens deze beoordelaar fungeren eerder gelezen annotaties als vergelijkingsmateriaal voor de andere annotaties die op hetzelfde onderwerp betrekking hebben. Annotaties die vervolgens gelezen worden, worden dan sneller gezien als aftreksels van de eerdere annotatie, aldus de beoordelaar.

3 Inventarisatie mogelijke kwaliteitsindicatoren

De volgende stap was om de elementen te achterhalen waarop de beoordelaars hun kwaliteitsoordelen hadden gebaseerd en daarmee criteria te achterhalen aan de hand waarvan de annotaties zijn (en worden) beoordeeld. Dat het achterhalen van kwaliteitsindicatoren niet eenvoudig zou zijn, laat de beperkte mate van overeenstemming zien. Deze suggereert dat beoordelaars mogelijk verschillende criteria aanleggen om annotaties te beoordelen.

De moeilijkheid om kwaliteitsindicatoren te achterhalen bleek voorts uit oriënterende gesprekken die zijn gehouden. In de beginfase van het onderzoek is op verschillende momenten aan collega's van verschillende faculteiten gevraagd wat een annotatie tot een goede annotatie maakt. De antwoorden varieerden, maar kwamen meestal overeen in vaagheid ('een goede annotatie moet me aan het denken zetten', 'een noot moet me iets leren').

Dat de kans dat beoordelaars aangeven dezelfde criteria te hanteren gering is, neemt niet weg dat voor iedere beoordelaar afzonderlijk kan worden achterhaald welke criteria zij hebben aangelegd om de annotaties te beoordelen. Om die criteria scherp te krijgen, zijn eerst mogelijke criteria opgesteld waarmee de beoordelaars de annotatie zouden hebben kunnen beoordeeld. De criteria zijn opgesteld door literatuur over annotaties te bestuderen, door informatie op websites van *law reviews* te achterhalen met daarop een instructie voor auteurs over hoe annotaties (*case notes*) te schrijven²⁸ en door oriënterende gesprekken te voeren met gepromoveerde juristen, verbonden aan een academische instelling, over wat een annotatie tot een goede annotatie maakt. Verder is gebruik gemaakt

²⁸ Een instructie is gevonden op de website van Montana-Law-Review (2008). 'Suggested Structure of a Law Review Note.' Retrieved March 17, 2011, from <http://www.montanalawreview.com/sitebuildercontent/sitebuilderfiles/casenoteoutline.doc>.

van een aantal toegestuurde juryrapporten.²⁹ Deze rapporten gaan over de verkiezing van de beste annotatie in een bepaald jaar (2006 respectievelijk 2007), gepubliceerd in *AB*. Het aardige is niet alleen dat de jury in kwestie twee keer een jaargang annotaties heeft bestudeerd en dus een overzicht heeft van een substantieel aantal annotaties, maar ook dat de jury heeft getracht de criteria te expliciteren aan de hand waarvan de annotaties zijn beoordeeld.

De criteria die in de twee rapporten worden genoemd, blijken uiteen te lopen. Zo worden in het rapport van 2006 als criteria genoemd: het plaatsen van de annotatie in een helder, meer algemeen verband, het leggen van verbanden met andere rechtspraak, wetgeving of relevante leerstukken (die er niet met de haren bijgesleept zijn) en het geven van een opinie.³⁰ In het juryrapport van 2007 wordt genoemd: een juist begrip van de uitspraak, het verklaren van de uitspraak binnen het systeem van het recht en het desbetreffende leerstuk, deugdelijkheid van het bronnenonderzoek, de diepgang van de analyse en het commentaar, wat uitgelegd wordt als het bespreken van de achtergrond van de betrokken waarden en het zich bewust zijn van eigen vooronderstellingen en, tot slot, het doorgronden van de betekenis van de uitspraak voor toekomstige rechtsontwikkeling en/of de uitvoeringspraktijk.

De hiervoor genoemde criteria zijn uitgewerkt in een vragenlijst met bijbehorende toelichting. De vragenlijst en toelichting zijn gebruikt om voor een hoeveelheid annotaties na te gaan of, en zo ja, welke van deze criteria te relateren zijn aan de kwaliteit van annotaties. Tevens is geanalyseerd in hoeverre verschillende beoordelaars dezelfde criteria aanleggen. Criteria die heel algemeen zijn geformuleerd (bijvoorbeeld originaliteit, betrouwbaarheid, creativiteit) zijn niet als zodanig meegenomen, ten eerste vanwege de algemeenheid en ten tweede omdat zij waarschijnlijk opgaan in andere, meer concrete factoren.

Beschrijven/interpreteren, verduidelijken, ordenen, doorrekenen

In de literatuur worden verschillende kenmerken van annotaties genoemd. Een eerste kenmerk is het *beschrijven en/of interpreteren* van de uitspraak waar de annotatie bij is geschreven. Dit gebeurt in veel, zo niet alle annotaties.³¹ Volgens Bloembergen is een goede interpretatie van een arrest essentieel voor een goede noot. Volgens hem kan uitleggen en analyseren alleen als de uitspraak goed

²⁹ Handleiding/Juryrapport AB 2007/2008 (niet gepubliceerd). Het gaat hier om een interne handleiding van een tijdschrift waarin annotaties worden gepubliceerd en waar de juryrapporten als bijlagen zijn bijgesloten.

³⁰ Ook het niet willen maken van een punt dat slechts zijdelings aan de orde is in de uitspraak wordt als criterium genoemd, maar hierbij wordt opgemerkt dat dit niet vaak voorkomt.

³¹ Montana-Law-Review (2008). 'Suggested Structure of a Law Review Note.' Retrieved March 17, 2011, from <http://www.montanalawreview.com/sitebuildercontent/sitebuilderfiles/casenoteoutline.doc>; Vranken 2005; Snijders 2003; Bloembergen 2002.

wordt gelezen (soms wel tien keer). Verder moet de annotator volgens hem thuis zijn in cassatietechniek. Hij moet, aldus Bloembergen, onder andere weten wat het onderscheid is tussen feit en recht, wat een feitelijke grondslag is en welke eisen aan de motivering van uitspraken worden gesteld. Voorts is het kennen van het spraakgebruik van de Hoge Raad van belang, maar dan nog blijkt het moeilijk de betekenis en draagwijdte in te schatten. Soms zelfs vinden wetenschappers van elkaar, ook gezaghebbende, dat zij verkeerde conclusies aan de uitspraak verbinden. Voor een deel – met name bij de uitleg en interpretatie van uitspraken – heeft deze verwarring volgens Bloembergen te maken met de verborgen ontstaansgeschiedenis van de uitspraak: de raadkamerdiscussie, concepten en notawisselingen worden niet naar buiten gebracht, terwijl een officiële toelichting op de uitspraak ontbreekt.

Behalve beschrijven/interpreteren worden ook het *verduidelijken* (het uitleggen van het wettelijke systeem en de achtergronden), *ordenen/systematiseren* (het vergelijken/inpassen van de uitspraak met/in eerdere uitspraken) en het *doorrekenen* (het analyseren van (mogelijke) consequenties voor soortgelijke gevallen) genoemd als kenmerken van annotaties.³²

Bronnengebruik

In het onderzoek van Crombag/Cohen 1974 is de hypothese geformuleerd dat de inhoud van een annotatie, en daarmee vermoedelijk ook de kwaliteit, afhankelijk is van de breedte van de literatuur waarover de annotator beschikt.³³ De gedachte hierachter is dat hoe meer associaties de annotator kan maken, des te breder zijn overzicht is en des te beter de annotator de uitspraak een plaats kan geven in het wettelijk systeem. Een breed overzicht maakt het voorts mogelijk een zo zuiver mogelijke voorspelling te geven van de gevolgen van de geannoteerde uitspraak. Dit voert naar het vermoeden dat de bronnen die de annotator gebruikt en de mate waarin hij dat doet, te relateren zijn aan de kwaliteit van de annotatie.

Relevantie

Voorts zou ook de relevantie van een annotatie de kwaliteit van annotaties kunnen voorspellen. Bij de relevantie kan worden onderscheiden tussen nieuws-waarde (is het probleem dat aan de orde is nieuw?), nut (is het probleem wezen-

³² Montana-Law-Review (2008). 'Suggested Structure of a Law Review Note.' Retrieved March 17, 2011, from <http://www.montanalawreview.com/sitebuildercontent/sitebuilderfiles/casenoteoutline.doc>. Vranken 2005; Snijders 2003; Handleiding/Juryrapport AB 2007/2008 (niet gepubliceerd). Zo ook Bleeker 2010.

³³ Crombag & Cohen 1974.

lijk of draagt de annotatie wezenlijk bij aan het oplossen van het probleem?) en reikwijdte (hoe groot is het probleem?).³⁴

Het eerder vermelde literatuur- en vooronderzoek liet zien dat annotaties ‘iets’ moeten toevoegen, de lezer aan het denken dienen te zetten of informatie moeten geven die de lezer nog niet had. Dit suggereert dat de relevantie van de annotatie ertoe doet. Hetzelfde geldt mogelijk voor de relevantie van de uitspraak. De eerder genoemde dienstbaarheid van juridisch onderzoek aan de rechtspraak³⁵ kan tot gevolg hebben dat de relevantie van de uitspraak doorwerkt in de relevantie van de annotatie.

4 Relatie mogelijke kwaliteitsindicatoren – beoordelingen van annotaties

Om na te gaan welke van de hiervoor genoemde criteria gerelateerd zijn aan de kwaliteit van annotaties, is in essentie de volgende procedure gevolgd. Drie beoordelaars hebben voor iedere van de aan hen toegewezen annotaties een vragenlijst ingevuld. Dit hield in dat zij een annotatie eerst van een oordeel dienden te voorzien. Het overige deel van de vragenlijst bestond uit items die mogelijke kwaliteitsindicatoren vertegenwoordigden.

Aan de hand van een statistische toets kon worden bekeken of de scores op de items uit de vragenlijst verschilden voor goede/uitstekende annotaties enerzijds en matige/slechte annotaties anderzijds. Een verschil in score op een item duidt op een relatie tussen het betreffende item uit de vragenlijst en de beoordeling (kwaliteit) van een annotatie.

Twee privaatrechtelijke beoordelaars en één strafrechtelijke beoordelaar hebben de vragenlijst onafhankelijk van elkaar ingevuld voor een hoeveelheid privaatrechtelijke respectievelijk strafrechtelijke annotaties. De beoordelingen (kwaliteitsoordelen) zijn gemeten op een vierpuntschaal (slecht – matig – goed – uitstekend). Ook de items uit de vragenlijst zijn gemeten met vierpuntschalen.

Omdat bij de privaatrechtelijke annotaties twee beoordelaars betrokken waren, kon worden berekend in hoeverre hun scores op de items overeenstemden. Dat is zinvol vanuit het oogpunt van betrouwbaarheid, aangezien op deze manier kan worden getoetst of twee beoordelaars die dezelfde annotaties lezen dezelfde scores toekennen waar het gaat om de aan- of afwezigheid van bepaalde kenmerken.

De mate van overeenstemming is getoetst door, voor de items uit de vragenlijst, de scores op de vierpuntschalen terug te brengen tot dichotome scores

³⁴ Tijssen 2009, p. 56, 59. Zie ook Montana-Law-Review (2008). ‘Suggested Structure of a Law Review Note.’ Retrieved March 17, 2011, from <http://www.montanalawreview.com/sitebuildercontent/sitebuilderfiles/casenoteoutline.doc>.

³⁵ Zie voetnoot 7.

(tweepuntschaal). De resultaten van deze betrouwbaarheidstoets zijn terug te vinden in Bijlage 1. Daaruit volgt een hoge mate van overeenstemming tussen de beoordelaars voor een behoorlijk aantal items. Voor een aantal items is de mate van overeenstemming laag.³⁶

Om per beoordelaar te achterhalen welke criteria zij hebben aangelegd om de annotaties te beoordelen, is de volgende procedure gevolgd. Het kwaliteitsoordeel (de beoordeling), dat met een vierpuntschaal is gemeten, is teruggebracht tot een tweepuntschaal (slecht/matig versus goed/uitstekend). Vervolgens is aan de hand van een serie T-toetsen gekeken of de scores op de items uit de vragenlijst verschilden tussen deze twee categorieën (slecht/matig versus goed/uitstekend). Met andere woorden, voor ieder item is bekeken of de annotaties die door de beoordelaars als goed/uitstekend zijn aangemerkt gemiddeld een andere score hadden dan de annotaties die als slecht/matig zijn aangemerkt. Deze vergelijking is gemaakt voor de beoordelaars afzonderlijk. Er bestond immers een reële kans dat zij verschillende criteria hanteerden om annotaties te beoordelen.

De beoordelaars hebben ieder zestien annotaties beoordeeld aan de hand van de vragenlijst. Hiervan is één privaatrechtelijke annotatie niet in de analyse meegenomen. Bij deze annotatie gaven zowel één beoordelaar als degene die de selectie maakte aan dat de annotatie meer het karakter van een artikel had dan van een annotatie. De andere privaatrechtelijke beoordelaar gaf desgevraagd aan ook twijfels te hebben gehad bij het aanmerken van de publicatie als annotatie.

Tabel 6 laat zien welke scores significant verschillen indien goede/uitstekende annotaties worden vergeleken met slechte/matige annotaties. Omwille van de overzichtelijkheid staan in de tabel alleen de significante resultaten weergegeven, dat wil zeggen de resultaten waarvan de kans klein is ($< 1\%$, $< 5\%$ of $< 10\%$) dat ze op toeval berusten. De volledige versie van deze tabel is te vinden in Bijlage 2.³⁷

³⁶ Het ontwikkelen van de vragenlijst heeft ongeveer een jaar geduurd. Tijdens het testen, zeker in de eindfase, bestond de indruk dat de vragen op dezelfde manier werden geïnterpreteerd. Het is daarom niet waarschijnlijk dat een gebrek aan overeenstemming tussen de beoordelaars voortvloeit uit een verschil in interpretatie.

³⁷ *Effect sizes* staan in diezelfde bijlage.

Tabel 6 – Verschillen scores tussen matige/slechte annotaties en goede/uitstekende annotaties

Criterium	Beoordelaar	Kwaliteit		Effect sizes (d)
		Matig/slecht	Goed/uitstekend	
Opbouw/toegankelijkheid	A (Pr)	2.89†	3.83†	.48
	C (Sr)	2.75†	3.17†	.37
Wet (interne rechtsvergelijking)	C (Sr)	.25†	.00†	.41
Wetsgeschiedenis (externe rechtsvergelijking)	B (Pr)	.00†	.25†	.11
Jurisprudentie (binnen rechtsgebied)	C (Sr)	1.25*	2.58*	.65
Jurisprudentie (externe rechtsvergelijking)	B (Pr)	.00†	.50†	.38
Literatuur (binnen rechtsgebied)	A (Pr)	1.00*	1.83*	.67
Gebrek aan bronnen	C (Sr)	3.00***	1.33***	.996
Verduidelijken (voldoende uitgebreid)	B (Pr)	2.00*	3.50*	.73
	C (Sr)	2.00*	3.58*	.70
Verduidelijken (voldoende diepgaand)	B (Pr)	2.36*	3.25*	.60
	C (Sr)	1.25***	3.50***	1.00
Ordenen (voldoende uitgebreid)	C (Sr)	.25*	2.50*	.67
Ordenen (voldoende diepgaand)	C (Sr)	.25*	2.33*	.70
Toetsen (overig) (voldoende diepgaand)	B (Pr)	.27†	1.25†	.39
Essentiële vragen/aspecten gedistilleerd	C (Sr)	2.00***	4.00***	.47
Uitspraak nieuw	B (Pr)	2.27†	3.25†	.998
Reikwijdte uitspraak	B (Pr)	2.27*	3.25*	.46
Signaleren vraag-/knelpunten	B (Pr)	1.73†	2.50†	.49
	C (Sr)	1.50*	2.92*	.70
Onderbouwd standpunt	B (Pr)	2.14*	3.00*	.58
	C (Sr)	1.00*	2.89*	.52
Praktische waarde annotatie	B (Pr)	2.18*	3.33*	.61
	C (Sr)	1.50**	2.60**	.89

Noot: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$. De tabel toont alleen significante verschillen en trends. Indien een antwoord ontbrak of de vraag niet van toepassing was, zijn deze gegevens niet in de analyse meegenomen. Hierdoor kan het aantal annotaties dat per item in de analyse is betrokken, verschillen.

Box 2 – Significantie

De kwaliteitsoordelen (beoordelingen) zijn gemeten met een vierpuntschaal, die is teruggebracht tot een dichotome variabele (tweepuntschaal). De items uit de vragenlijst zijn gemeten met een vierpuntschaal. Een verschil tussen bijvoorbeeld 2.27 (slecht/matig) en 3.25 (goed/uitstekend) betekent dat in de slechte/matige annotaties de score op het betreffende item bijna een punt lager ligt dan bij de goede/uitstekende annotaties. Hoe groter het verschil, hoe groter de rol kan zijn van het betreffende item. Verschillen zijn statistisch significant, dat wil zeggen dat de kans klein is dat zij op toeval berusten, indien $p < .05$. Hoe lager de p -waarde, hoe kleiner de kans dat de gevonden verschillen op toeval berusten. Bij $p < .10$ zijn de verschillen slechts indicatief. Tot slot geldt voor alle van de gevonden verbanden dat het gaat om correlaties en niet om causale relaties.

Effect sizes ('d') geven de sterkte van het effect aan. *Effect sizes* van ongeveer 0.2 of 0.3 worden als klein aangemerkt, van 0.5 als matig en van 0.8 of hoger als sterk tot zeer sterk.

Hoewel het aantal annotaties dat is onderzocht, gering is, laat de tabel toch een hoeveelheid significante resultaten zien. Uit de tabel volgt dat er soms overeenkomsten en soms verschillen tussen beoordelaars bestaan ten aanzien van de criteria waarmee goede/uitstekende annotaties van matige/slechte annotaties kunnen worden onderscheiden. De opbouw en toegankelijkheid lijken voor beoordelaar A (privaatrecht) en beoordelaar C (strafrecht) relevant te zijn om goede/uitstekende annotaties van matige/slechte annotaties te onderscheiden, hoewel het gevonden effect statistisch gezien slechts indicatief is ($p < .10$).

Tussen beoordelaar B en C is meer overeenstemming. Het verduidelijken, signaleren van vraag- en/of knelpunten, het geven van een onderbouwd standpunt en de praktische waarde van de annotatie blijken voor deze beoordelaars (mede)beslissend te zijn om een annotatie als goed/uitstekend aan te merken of juist als matig/slecht. Bij één van de drie beoordelaars zijn voorts verschillen gevonden bij een aantal items, maar deze verschillen zijn hoogstens indicatief ($p < .10$). Het gaat om de bronnen wet (interne rechtsvergelijking), wetsgeschiedenis (externe rechtsvergelijking), jurisprudentie (binnen rechtsgebied), jurisprudentie (externe rechtsvergelijking) en literatuur (binnen rechtsgebied), een gebrek aan bronnen, het ordenen en toetsen, het distilleren van essentiële vragen/aspecten uit de uitspraak, de vraag of de uitspraak nieuw is en om de reikwijdte van de uitspraak. Er zijn geen criteria gevonden die alle drie de beoordelaars zeggen aan te leggen om goede/uitstekende van matige/slechte annotaties te onderscheiden.

De gevonden effecten zijn zwak tot matig ($d < .50$) tot groot ($d > .75$), zie nader Bijlage 2. Dat valt tevens op wanneer wordt gekeken naar de verschillen.

Deze variëren tussen de 0.8 punt en de 1.7 punten, wat vrij behoorlijk is op een vier- of vijfpuntschaal.

In het algemeen valt op dat de beoordelaars verschillende criteria aanleggen om een annotatie te beoordelen. Op zichzelf verbaast dit niet, maar de verwachting was dat een aantal factoren gemeenschappelijk zou zijn. Het idee was bijvoorbeeld dat beoordelaars met name het adequaat ordenen en doorrekenen in verband zouden brengen met de kwaliteit van annotaties. Dit is ook het beeld dat in de literatuur naar voren komt, maar dat wordt niet bevestigd door de resultaten van de uitgevoerde analyse.

Voor items over bronnengebruik zijn verschillen gevonden, maar uit de scores blijkt dat er relatief weinig naar bronnen wordt verwezen. Annotaties bevatten relatief weinig verwijzingen naar relevante regelgeving, wetsgeschiedenis en naar (rechtsvergelijkende) literatuur en jurisprudentie.³⁸ Dit lijkt kenmerkend voor annotaties. Meerdere beoordelaars merkten in dit verband dan ook op bij annotaties die relatief veel verwijzingen bevatten, dat deze soms ook voor een artikel kunnen doorgaan.

Daarnaast laten de resultaten zien dat annotaties over het algemeen ‘volgzaam’ zijn. Zij stellen over het algemeen het oordeel van de rechter centraal. In weinig annotaties behandelt de annotatie andere aspecten of probleempunten dan die in de uitspraak aan de orde zijn. Dat gebeurt slechts in twee van de vijftien annotaties (volgens beoordelaar A, privaatrecht), vijf van de vijftien annotaties (volgens beoordelaar B, privaatrecht) respectievelijk in twee van de zestien annotaties (volgens beoordelaar C, strafrecht). Daarbij geldt dat als er een nieuw probleem of aspect aan de orde wordt gesteld, dit geen wezenlijk nieuw punt of probleem hoeft te zijn, zo laten de antwoorden van de beoordelaars zien. Ook het leveren van een constructieve bijdrage aan het oplossen van een geconstateerd probleem, het onderbouwen van eigen standpunten, het bespreken van voorstellen of standpunten van anderen alsook het weerleggen van die standpunten gebeurt in circa de helft van de annotaties, zo constateren de beoordelaars, maar ook daarbij geldt dat als de annotator er aandacht aan besteedt, deze aandacht beperkt is.

Oorspronkelijk was het de bedoeling dat voor ieder van de drie rechtsgebieden twee deskundige beoordelaars onafhankelijk van elkaar de vragenlijst zouden invullen voor een reeks annotaties. Na drie beoordelaars bestond echter het beeld en de verwachting dat een uitbreiding van het onderzoek louter de conclusie zou bevestigen dat beoordelaars verschillende criteria aanleggen om de kwaliteit van annotaties te beoordelen. Dit, in combinatie met de omstandigheid dat het beoordelen van de annotaties aan de hand van de opgestelde vragenlijst een tijdrovende klus is voor een beoordelaar (ca. 30-45 minuten per annota-

³⁸ Met name beoordelaar 1 vond het belangrijk dat literatuur werd behandeld in een annotatie. Dit criterium bleek echter niet (mede)bepalend te zijn voor het kwaliteitsoordeel.

tie), is besloten om het hierbij te laten en dit deel van het onderzoek niet verder voort te zetten.

5 Conclusie

Annotaties, ook de goede/uitstekende, stellen de uitspraak als zodanig in beperkte mate ter discussie. Annotatoren analyseren of bekritisieren hoe de uitspraak zich, gegeven die uitspraak, verhoudt tot eerdere uitspraken of bepaalde beginselen, of wat de gevolgen ervan zijn voor de wetsystematiek en de praktijk. Hieruit blijkt de volgzzaamheid.

Waarschijnlijk onderscheiden annotaties zich op dit punt, samen met het bronnengebruik, van artikelen. Gevraagd naar het onderscheid tussen een annotatie en een artikel, gaven twee beoordelaars aan dat dit vooral te maken had met – inderdaad – het volgen van de uitspraak en het beperkte bronnengebruik. In tegenstelling tot een artikel of een dissertatie, waarin veel meer de problematiek centraal staat in plaats van de uitspraak en de daarin gewezen rechtsregel, kan binnen een annotatie worden volstaan met het bespreken van de gevolgen van de uitspraken en de houdbaarheid ervan in het licht van het geldende recht. Bij een annotatie die verder gaat, wordt al snel de vraag gesteld of het dan nog wel gaat om een annotatie, of om een artikel, zo gaven beoordelaars tijdens dit deel van het onderzoek aan. Dit bevestigt het beeld dat annotaties zich voornamelijk concentreren op de gewezen uitspraak.

Een belangrijke uitkomst is dat er een matige tot redelijk grote mate van overeenstemming bestaat over de kwaliteit van annotaties, maar dat de beoordelingen niet eenduidig zijn. De resultaten laten zien dat er tussen de 50% en 86% overeenstemming bestaat over de kwalificatie van de geanalyseerde annotaties als goed/uitstekend en matig/slecht. Een herhaalde meting wekt de indruk dat beoordelaars anders kunnen oordelen over een annotatie naarmate zij de annotatie vaker analyseren of beoordelen, hoewel enige terughoudendheid is geboden, aangezien de conclusie is gebaseerd op een relatief kleine hoeveelheid gevallen.

Uit de statistische toetsen die zijn uitgevoerd, volgt verder dat beoordelaars verschillende redenen aanvoeren om hun beoordelingen te motiveren. Tegelijk komt uit de analyse een aantal aspecten naar voren dat zich bij goede of uitstekende annotaties vaker of in belangrijkere mate manifesteert dan bij minder geslaagde annotaties. Deze aspecten zijn onder te brengen in, wat zou kunnen worden aangeduid als, structuuraspecten, verantwoord bronnengebruik, praktische relevantie en evaluatieve kenmerken.

Ten aanzien van de mate van overeenstemming kan worden vastgesteld dat deze soms matig is, soms hoog, maar zeker niet dermate laag dat sprake is van willekeurigheid. Dit roept de vraag op of de beoordeling niet meer te uniformeren zou zijn, bijvoorbeeld door beoordelaars specifieke instructies te geven

over wat de kwaliteit van annotaties bepaalt. Een dergelijke instructie ontbrak bij het invullen van de vragenlijst. In het uitgevoerde onderzoek werden de beoordelaars louter geïnstrueerd over welke vragen zij moesten beantwoorden en hoe deze vragen moesten worden opgevat. De beoordelaars kregen echter geen informatie of opdrachten over wanneer een annotatie als goed of uitstekend mag worden gekwalificeerd; daar waren de beoordelaars vrij in.

Resteert nog de vraag hoe moet worden verklaard dat de beoordelaars verschillende criteria stellen aan te leggen voor het beoordelen van annotaties. Een eerste verklaring is dat de vragenlijst niet deugt. Dit ligt niet voor de hand, aangezien de vragenlijst zorgvuldig is samengesteld op basis van bestaande literatuur en oriënterende gesprekken met deskundigen die ervaring hebben met het schrijven van annotaties en met het lezen ervan. Daarbij komt dat uitvoerig – ruim een jaar – met de vragenlijst is proefgedraaid. Met twee deskundigen zijn zestien annotaties soms zelfs meerdere keren doorgenomen. In de eindfase bestond, ook bij de beoordelaars zelf, de indruk dat de vragen op dezelfde wijze werden geïnterpreteerd.

Een tweede mogelijke verklaring voor het ontbreken van algemene criteria is dat beoordelaars verschillende criteria aanleggen om annotaties te beoordelen. Een derde verklaring, die ook niet kan worden uitgesloten op basis van de gevonden resultaten, is dat beoordelaars zelf niet goed kunnen benoemen waarom bepaalde annotaties kwalitatief beter zijn dan andere annotaties. Op basis van de analyses die in dit hoofdstuk zijn uitgevoerd, is geen uitsluitsel te geven over welke van deze twee verklaringen het meest overtuigt. Meer hierover in Hoofdstuk 3.

Kwaliteitsindicatoren blijkend uit annotaties

1 Inleiding en methode

Uit de analyse die is uitgevoerd in Hoofdstuk 2 schreven de beoordelaars de kwaliteit van annotaties toe aan uiteenlopende factoren. Onduidelijk was echter of de verschillen in kwaliteitsoordelen moesten worden verklaard doordat beoordelaars verschillende criteria aanleggen om annotaties te beoordelen, of doordat beoordelaars zelf niet goed in staat waren om te benoemen waarom bepaalde annotaties kwalitatief beter zijn dan andere annotaties.

Om hier meer zicht op te krijgen, is in dit hoofdstuk voor een andere benadering gekozen. In plaats van aan beoordelaars te vragen om een annotatie te scoren aan de hand van vooraf opgestelde criteria, zijn in dit hoofdstuk de kwaliteitsindicatoren vastgesteld op basis van een analyse van de annotaties zelf. Door te onderzoeken of de kwaliteitsoordelen van annotaties kunnen worden verklaard door de aan- of afwezigheid van bepaalde kenmerken, kunnen mogelijke indicatoren worden opgespoord. Het voordeel van deze benadering ten opzichte van die uit het vorige hoofdstuk is dat kan worden uitgesloten dat de kenmerken die de kwaliteit van annotaties zouden kunnen verklaren, voortkomen uit informatie van beoordelaars die niet goed weten welke criteria zij aanleggen of door beoordelaars die zeggen verschillende criteria aan te leggen, terwijl niet duidelijk is of dit het geval is.

Het beantwoorden van de vraag wat betrouwbare indicatoren zijn om de kwaliteit van annotaties te beoordelen, gebeurt door te onderzoeken welke kenmerken annotaties hebben die de beoordelaars als goed/uitstekend hebben aangemerkt in vergelijking met de kenmerken van annotaties die als matig/slecht zijn aangemerkt. Indien de twee groepen annotaties (goed/uitstekend versus matig/slecht) andere kenmerken hebben, kunnen hieruit kwaliteitsindicatoren worden afgeleid. Hetzelfde geldt als bepaalde kenmerken alleen bij annotaties worden aangetroffen die als goed/uitstekend zijn aangemerkt: deze kenmerken zijn dan als mogelijke kwaliteitsindicatoren te beschouwen.

Voor de analyse kwamen in beginsel alle 18 bestuursrechtelijke, 16 privaatrechtelijke en 16 strafrechtelijke annotaties in aanmerking die door de deskundigen uit het betreffende rechtsgebied zijn geselecteerd. Om dezelfde redenen als in het vorige hoofdstuk is een aantal annotaties niet meegenomen in de analyses. Het gaat om drie bestuursrechtelijke annotaties en één privaatrechtelijke annotatie. Dit brengt het totaal aantal bestudeerde annotaties op 46.

De werkwijze die is gevolgd om vast te stellen welke kenmerken een bepaalde annotatie heeft, is dat steeds in de kantlijn van de annotaties passages zijn gecodeerd met codes die bepaalde kenmerken vertegenwoordigen. Het coderen is voornamelijk globaal gedaan, dat wil zeggen per alinea, paragraaf of nummer (niet per zin).

Gekozen is voor een iteratief proces. Dit houdt in dat de kenmerken gaandeweg de analyses zijn opgesteld, aangevuld en aangescherpt. Niet gekozen is voor een werkwijze waarbij de kenmerken vooraf zijn geformuleerd. De reden hiervoor is dat de vooraf geformuleerde kenmerken mogelijk niet geschikt zijn om annotaties mee te beoordelen en dus geen kwaliteitsindicatoren zijn. Het gevaar met het werken van vooraf geformuleerde kenmerken is dat deze kenmerken, zoals in voorgaand hoofdstuk, worden opgesteld aan de hand van wat bekend is binnen het juridisch-academisch forum, terwijl niet zeker is dat dit forum een scherp beeld heeft van wat een annotatie tot een goede annotatie maakt.

Om te voorkomen dat voortschrijdend inzicht tot andere coderingen zou leiden dan die in eerste instantie zijn aangebracht, zijn de annotaties meerdere malen bekeken en bestudeerd. Om na te gaan of het overzicht van kenmerken betrouwbaar was, dat wil zeggen, te controleren dat anderen de annotaties vergelijkbaar zouden coderen, is ongeveer de helft van de zestien privaatrechtelijke annotaties samen met een student-assistent gecodeerd. Deze student-assistent heeft eerst zelf de annotaties gecodeerd, waarna de coderingen met elkaar zijn vergeleken. Soms leidde dat tot aanpassingen van de coderingen, maar over het algemeen was er een redelijke tot grote mate van overeenstemming. Gezien de relatief grote hoeveelheid kenmerken was het niet mogelijk deze overeenstemming in percentages of statistische waardes (kappa) uit te drukken.

Na het coderen van de annotaties is bekeken welke annotaties als goed/uitstekend zijn aangemerkt, welke als matig/slecht en, vervolgens, of het verschil in waardering kan worden verklaard door de aan- of afwezigheid van bepaalde kenmerken. Daartoe is per rechtsgebied een matrix opgesteld waaruit deze verbanden zouden moeten blijken. De matrices zijn opgenomen in Bijlage 4.

De analyse is uitgevoerd voor ieder rechtsgebied (bestuursrecht, privaatrecht, strafrecht) afzonderlijk. Dit voorkomt dat mogelijke verschillen tussen rechtsgebieden over het hoofd worden gezien. Dit hoofdstuk geeft per rechtsgebied de belangrijkste bevindingen weer. Een uitgebreidere analyse is te vinden in Bijlage 4. Het slot van dit hoofdstuk bevat een overzicht waarin de observaties van de analyse van de drie rechtsgebieden zijn samengevoegd. Aan de hand daarvan is bekeken of er gemeenschappelijke conclusies kunnen worden getrokken en of er bepaalde indicatoren zijn die specifiek gelden voor een bepaald rechtsgebied.

Omdat logischerwijs niet alleen de aan- of afwezigheid van bepaalde kenmerken bepalend zou zijn voor de kwaliteit van annotaties, maar ook in welke mate bepaalde kenmerken terug zouden komen in annotaties, zijn de annotaties steeds nader bekeken om te bezien op welke wijze en in welke mate de verschillende kenmerken zijn uitgewerkt in de verschillende annotaties. De verslaglegging van deze analyse volgt hierna.

De beschrijvingen en illustraties van de kenmerken en de annotaties die hierna volgen, zijn algemeen van aard. De reden hiervoor is dat aan de beoordelaars is beloofd dat oordelen of uitlatingen niet tot de beoordelaar herleidbaar zijn. Ook omdat het niet het doel is van dit onderzoek om bepaalde annotaties of annotatoren in een goed of kwaad daglicht te stellen, zijn de annotaties in meer algemene zin omschreven. Om die reden worden de annotaties ook aangeduid met nummers (in plaats van vindplaatsen). Om toch inzicht te geven in hoe de verschillende kenmerken terugkomen in de verschillende annotaties, zijn korte omschrijvingen van de annotaties opgenomen in Bijlage 3.

2 Kenmerken van annotaties

Bij het bestuderen van de annotaties viel al vrij snel op dat de kenmerken, zowel qua hoeveelheid als qua omschrijving, afwijken van wat in de literatuur te lezen valt, in die zin dat de in de literatuur omschreven kenmerken algemener zijn en soms meerdere van de volgende kenmerken omvat. De kenmerken die zijn aangetroffen, kunnen in essentie worden onderverdeeld in beschrijvende en evaluatieve kenmerken. Zij worden hierna nader omschreven.

Beschrijvende kenmerken:

- *Hoogste rechter*. Dit kenmerk gaat over welke rechter de uitspraak heeft gewezen waarbij de annotatie is geschreven (rechtbank, hof, hoogste rechter, nationaal, Europees, internationaal).
- *Beschrijving/samenvatting casus en/of uitspraak*. De annotator geeft de feiten en/of de uitkomst (de beslissing) weer dan wel vat deze samen. Hieronder valt ook het toelichten, interpreteren of uitleggen van passages, tenzij de annotator constateert dat er sprake is van een onduidelijkheid of van een motiveringsgebrek: in dat geval is sprake van ‘Onduidelijke passages constateren c.q. ophelderen / motiveringsgebrek opmerken’.
- *Beschrijven geldend recht/verwijzen naar literatuur of jurisprudentie daarover*. Hierbij gaat het om een bespreking van het geldende recht dat op de uitspraak van toepassing is. Verwijzingen naar literatuur over (achtergronden van) het geldende recht vallen tevens onder dit kenmerk. Ook indien de literatuur kritisch is over het geldende recht, of over de toepassing ervan, is sprake van dit kenmerk, tenzij de annotatie verschillende theorieën of opvat-

tingen bespreekt en daarna nagaat bij welke theorie de uitspraak het beste aansluit (zie 'Nagaan welke leer is toegepast (welke leer geldend recht is)').

- *Wijzen op belang uitspraak voor rechtspraktijk of -theorie.* De annotator wijst op het belang van de uitspraak. Het gaat hierbij om een specifieke omschrijving van het belang (bijvoorbeeld: 'de uitspraak is relevant voor praktijkjuristen die deze grond inroepen') en niet om algemene bewoordingen ('deze interessante uitspraak gaat over ...'). Wat ook niet onder dit kenmerk valt, is de situatie waarin de annotator het belang van de uitspraak omschrijft in termen van een evaluatie ('de uitspraak is onduidelijk/deugt niet/gaat niet zonder kritiek, etc.').
- *Schetsen van kwesties die zich geregeld voordoen.* Dit kenmerk houdt in dat de annotator omschrijft welke vragen of problemen zich met enige regelmaat voordoen, bijvoorbeeld in de rechtspraktijk (bijvoorbeeld: 'de rechter ziet zich nogal eens geconfronteerd met de vraag of ...', 'advocaten lopen vaak aan tegen het probleem dat...').

Evaluatieve kenmerken:

- *Opwerpen vragen (zonder deze te beantwoorden).* De annotatie werpt één of meerdere vragen op ('hoe zit het met ...?', 'kan het zo zijn dat ...?') zonder deze werkelijk te beantwoorden. Het kan hierbij gaan om retorische vragen.
- *Nagaan welke leer is toegepast (welke leer geldend recht is).* Hiervan is sprake indien een annotator aan de hand van de uitspraak nagaat welke opvatting, benadering of leer (bijvoorbeeld de finale of intersubjectieve leer in het strafrecht bij het bepalen van opzet) is toegepast door de rechter. Vereist is dat in de literatuur (of rechtspraak) daadwerkelijk verschillende opvattingen of benaderingen worden aangehangen.
- *Criterium/criteria ontwikkelen en toetsen of de uitspraak aan dit criterium (deze criteria) voldoet.* De annotator ontwikkelt een criterium om gevallen te beoordelen zoals aan de orde in de uitspraak, van welk criterium hij of zij vervolgens nagaat of het wordt gebruikt door de rechter.
- *Implicaties van de uitspraak bespreken (bv. voor toepassing/interpretatie geldend recht).* De annotator gaat in op de gevolgen van de uitspraak voor de toepassing of interpretatie van het recht. Te denken valt aan de situatie waarin de annotator de gewezen rechtsregel toepast op nieuwe, andere gevallen, of aan het geval dat de annotator bespreekt wat de betekenis van de uitspraak is voor andere leerstukken of rechtsgebieden. Mogelijk, maar niet noodzakelijk, concludeert de annotator aan de hand hiervan dat de regel al dan niet bruikbaar, redelijk of anderszins wenselijk is. Het constateren van onduidelijkheden valt tevens onder dit kenmerk, mits wordt toegelicht wat de gevolgen zijn van de onduidelijkheid voor toekomstige gevallen. In an-

dere gevallen is sprake van ‘Onduidelijke passages constateren c.q. ophelderen/motiveringsgebrek opmerken’ (zie hierna).

- *Opvatting van andere auteur(s) toetsen/ter discussie stellen.* De annotator gaat aan de hand van de uitspraak na of de rechter een opvatting of benadering volgt die in de literatuur is voorgesteld of wordt aangehangen. Een andere mogelijkheid is dat de annotator een bepaalde opvatting of benadering, zoals die aan de orde is in de literatuur, ter discussie stelt, bijvoorbeeld door er kritische kanttekeningen bij te plaatsen of door haar te verwerpen. In tegenstelling tot het kenmerk ‘nagaan welke leer is toegepast’, hoeft bij dit kenmerk geen sprake te zijn van concurrerende opvattingen. Voldoende is dat kritiek wordt geuit op opvattingen of benaderingen van anderen.
- *Toetsen of uitspraak aansluit bij eerdere uitspraken.* De annotatie vergelijkt de besproken uitspraak met eerder gewezen uitspraken en gaat na of de besproken uitspraak afwijkt van eerdere jurisprudentie.
- *Toetsen of beslissing van de rechter beantwoordt aan doel van de regeling.* De annotatie beschrijft welk doel de regel of regeling nastreeft en of de uitspraak in overeenstemming is met dit doel.
- *Toetsen van de uitspraak aan 'hogere' recht (beginselen/verdragsbepalingen/EHRM-rechtspraak).* De annotator toetst of de uitkomst en/of de gewezen of de gebruikte rechtsregel in overeenstemming is met rechtsbeginselen, verdragsbepalingen, rechtspraak van Europees recht of andere internationale rechtspraak. Een vergelijking zonder evaluatie (‘in het verdrag ... wordt de problematiek op deze wijze opgelost’) valt niet onder het toetsen aan ‘hogere’ recht, maar onder beschrijving van het geldende recht.
- *Evalueren beslissing aan de hand van wat in de praktijk (van procesvoering) gebeurt/gangbaar is/gebruikelijk is.* De annotatie beschrijft een bepaalde praktijk (bijvoorbeeld het procesgedrag van een procederende partij) en toetst of de uitspraak in overeenstemming is met die praktijk.
- *Evaluatie van de uitspraak door andere denkbare uitkomsten af te zetten tegen de gekozen uitspraak.* De annotator gaat na wat er zou gebeuren als de rechter een andere beslissing zou hebben genomen (‘stel dat de rechter anders had besloten’) en gaat na of de beslissing die de rechter heeft genomen, de juiste is.
- *Systematisch de argumenten voor en tegen de beslissing nalopen.* De annotator inventariseert op systematische wijze welke argumenten voor en tegen de beslissing zijn aan te voeren en gaat in op de overtuigingskracht van de verschillende argumenten.
- *Annotator geeft (onderbouwd) aan of hij/zij het eens dan wel oneens is met de uitspraak.* De annotator beoordeelt of hij in de besproken uitspraak dezelfde beslissing zou hebben genomen. Hieronder valt ook het plaatsen van kritische kanttekeningen bij de individuele beslissing. De evaluatie en con-

clusies van de annotator blijven beperkt tot de individuele uitspraak en vallen niet onder een van de andere categorieën (bv. het bespreken van implicaties van de uitspraak).

- *Annotator zoekt naar een verklaring voor de uitkomst of de gewezen rechtsregel.* De annotator gaat op zoek naar de reden waarom de rechter een bepaalde beslissing heeft genomen ('wat is de verklaring voor ...?', 'Die keuze is wellicht mede ingegeven door ...').
- *Onduidelijke passages constateren c.q. ophelderen/motiveringsgebrek opmerken.* Dit kenmerk gaat over het opmerken van passages die bijvoorbeeld multi-interpretabel zijn, onzorgvuldig zijn geformuleerd, of tegenstrijdigheden bevatten. Ook kan sprake zijn van een gebrekkige motivering, in die zin dat de rechter volgens de annotator uitgebreider of beter had moeten uitleggen waarom zij tot een bepaalde beslissing is gekomen. De annotator kan onduidelijke passages of motiveringsgebreken opmerken en ophelderen, in die zin dat de annotator een poging doet meer duidelijkheid te creëren. Indien de annotator de gevolgen van de onduidelijkheid bespreekt, is sprake van 'implicaties van de uitspraak'.
- *Aangeven of een procespartij beter een andere keuze had kunnen maken in het proces (bv. andere rechtsgrond inroepen) c.q. wat (andere) mogelijkheden zijn om rechtsposities te versterken/wat er zou zijn gebeurd als een procespartij een andere keuze zou hebben gemaakt.* De annotator betoogt of beschrijft dat een van de procespartijen beter een andere keuze had kunnen maken, bijvoorbeeld door te stellen of te laten zien dat een procespartij beter een andere rechtsgrond aan had kunnen voeren of een andere juridische constructie had kunnen kiezen.
- *Kritiek op voorstellen vanuit de politiek.* De annotator uit kritiek op één of meerdere voorstellen die vanuit de politiek zijn gedaan.
- *Verwachting voor ontwikkeling van leerstuk.* De annotator schetst zijn of haar verwachting ten aanzien van de ontwikkeling van een bepaald leerstuk. Dit kenmerk wijkt af van het kenmerk 'schetsen van implicaties' doordat het schetsen van toekomstverwachtingen meer algemene verwachtingen betreft die niet direct uit de uitspraak hoeven te volgen (bijvoorbeeld 'de rechter zal in de toekomst omgaan', 'er zullen zich meer van dit soort zaken aandienen').

3 Observaties en voorlopige conclusies

Voor iedere annotatie is bekeken of de hiervoor genoemde kenmerken terug te vinden zijn in de annotaties die door de deskundigen zijn aangeleverd. Daarbij zijn drie scores toegekend, te weten '0' (kenmerk niet aangetroffen), '1' (kenmerk aangetroffen) en '0.5' (kenmerk aangetroffen maar zeer beperkt in omvang).

in vergelijking met andere annotaties waar het kenmerk is aangetroffen). De resultaten zijn voor ieder van de drie rechtsgebieden (bestuursrecht, privaatrecht, strafrecht) in een matrix geplaatst. Omdat aanwezigheid en lengte onvoldoende betrouwbaar zijn om de inhoud van annotaties te kunnen dekken, is voor iedere annotatie per kenmerk nader geanalyseerd hoe het terugkwam in de betreffende annotatie. De matrices en de resultaten zijn opgenomen in Bijlage 4. De informatie hierna beperkt zich om redenen van overzichtelijkheid tot de uitkomsten. Het is echter instructief om ook de analyse door te nemen die tot de hiernavolgende indrukken en voorspellingen heeft geleid.

3.1 Kwaliteitsindicatoren

Een eerste regelmatigheid die uit de drie rechtsgebieden af te leiden valt, is dat annotaties over het algemeen niet worden gekwalificeerd als goede of uitstekende annotaties als zij in essentie bestaan uit een weergave van de casus, de beslissing van de rechter en/of een beschrijving van het geldende recht dat op dergelijke gevallen van toepassing is. Indien een annotatie (vrijwel) uitsluitend hieruit bestaat, is dit dan ook een indicatie dat de academische kwaliteit van de annotatie matig of zelfs slecht is.

Een aantal bestuursrechtelijke annotaties met alleen dit kenmerk is toch als goed/uitstekend aangemerkt. Uit de toelichtingen van de beoordelaars volgde dat het bieden van een handzaam overzicht in een aantal gevallen wordt gewaardeerd. Het is echter niet mogelijk gebleken om criteria te achterhalen die zicht geven op wanneer het uitsluitend weergeven van feiten, de uitspraak en/of het geldende recht als goed/uitstekend moet worden aangemerkt.

Voorts kon voor een aantal andere kenmerken worden vastgesteld dat deze een positieve invloed hadden op de kwaliteit van annotaties. Deze positieve relatie blijkt echter alleen te bestaan indien de passages waarin deze kenmerken aan de orde zijn, een behoorlijke omvang hebben, dat wil zeggen dominant zijn binnen de betreffende annotatie, eventueel samen met het weergeven van feiten, de uitkomst en/of het geldende recht. Ook bij een beperkte omvang kan een annotatie positief worden gewaardeerd, maar dan is sprake van een combinatie met een van (onderstaande) kenmerken, en is de omvang van de passages samen van een behoorlijke, in ieder geval niet-geringe omvang.

Een van de kenmerken die positief is geassocieerd met de kwaliteit van annotaties, is het schetsen van implicaties. In alle drie de rechtsgebieden bleken de beoordelingen van annotaties te kunnen worden gerelateerd aan de aan- of afwezigheid van dit kenmerk. Een gebruikelijke wijze van het schetsen van implicaties is het aan de hand van andere, meestal nieuwe casus nagaan van wat de gevolgen van de uitspraak zijn en of deze gevolgen wenselijk zijn, eventueel in het licht van alternatieve oplossingen of uitkomsten.

Daarnaast kon worden vastgesteld dat het toetsen aan ‘hoger’ recht een positieve invloed heeft op de kwaliteit van annotaties. Het is echter niet mogelijk om op basis van de gedane analyse te bepalen of deze effecten voor alle drie de rechtsgebieden gelden. Het toetsen aan ‘hoger’ recht was slechts aan de orde in bestuursrechtelijke annotaties.

In de strafrechtelijke annotaties ontstond het beeld dat er een relatie bestond tussen de kwaliteit van annotaties en het nagaan welke leer was toegepast. De annotaties die met name bestaan uit het aan de hand van de uitspraak laten zien welke theorie of doctrine die in de literatuur is voorgesteld of wordt aangehangen het beste verklaart waarom de rechter in dit geval (en mogelijk in eerdere gevallen) tot een bepaalde beslissing is gekomen, zijn door beide beoordelaars als goed aangemerkt. De indruk bestond dat er een positieve relatie is tussen dit kenmerk en de kwaliteit van annotaties, maar enige twijfel is hier op zijn plaats, omdat ook andere factoren in het spel waren die de positieve relatie met de kwaliteit van annotaties mogelijk zouden kunnen verklaren. Overigens is dit kenmerk alleen teruggevonden in strafrechtelijke annotaties. Het is dan ook niet zeker dat dit kenmerk ook relevant is in andere rechtsgebieden. Dit zal waarschijnlijk afhankelijk zijn van de beschikbaarheid van theorieën en doctrines die met elkaar strijden om het beste verklaringsmodel te zijn, alsook van de omstandigheid in hoeverre de uitspraak laat doorschemeren welke leer zij aanhangt.

Uit de analyse van privaatrechtelijke annotaties volgt voorts dat het ontwikkelen van een criterium positief gerelateerd was aan de beoordeling van annotaties. Een vergelijkbare conclusie geldt voor het systematisch nalopen van voor- en tegenargumenten, inclusief het bespreken van de houdbaarheid van deze argumenten. Uit de analyse van strafrechtelijke annotaties kan worden opgemaakt dat dit kenmerk een indicatie is voor een goede of uitstekende annotatie indien het een centrale functie vervult in de betreffende annotatie.

3.2 Kenmerken met geen of zeer geringe invloed

Voor verschillende kenmerken kan op basis van de uitgevoerde analyse worden vermoed dat zij geen of een zeer geringe invloed hebben op de kwaliteit van annotaties. Dit geldt in de eerste plaats voor het vergelijken met eerdere jurisprudentie (in de zin van het vaststellen of de uitspraak wel/niet aansluit bij eerder gewezen uitspraken). In geen van de drie rechtsgebieden zijn aanwijzingen gevonden voor een verband met de kwaliteit van annotaties. Datzelfde geldt voor het wijzen op het belang van de uitspraak.

Voor een aantal kenmerken geldt dat het slechts is aangetroffen in annotaties afkomstig uit één rechtsgebied. De meeste van deze kenmerken bleken niet gerelateerd aan de kwaliteit van annotaties. Het gaat hierbij om het uiten van kritiek op voorstellen uit de politiek (bestuursrecht), het evalueren van een uit-

spraak aan de hand van wat gangbaar of gebruikelijk is in de praktijk (bestuursrecht), het schetsen van kwesties die zich geregeld voordoen (strafrecht), het toetsen of de beslissing beantwoordt aan het doel van de regeling (bestuursrecht), om het opwerpen van vragen (zonder deze te beantwoorden) (strafrecht) en om het laten zien dat procespartijen beter een andere keuze hadden kunnen maken (privaatrecht). Voor zover er een invloed is van het toetsen of ter discussie stellen van opvattingen van andere auteurs (privaatrecht), is die invloed beperkt. Zij lijkt afhankelijk te zijn van de aanwezigheid van andere kenmerken, zodat het mogelijk is dat de kwaliteit door die andere kenmerken wordt verklaard, of door een combinatie van kenmerken, waarbij moeilijk is vast te stellen wat de rol is van het toetsen of ter discussie stellen van opvattingen van andere auteurs.

Ten aanzien van het constateren c.q. ophelderen van onduidelijke passages en motiveringsgebreken, dat in annotaties binnen alle drie de rechtsgebieden terugkwam, zijn onvoldoende aanwijzingen gevonden om van een kwaliteitsindicator te kunnen spreken. Alleen bij de strafrechtelijke annotaties leek er een mogelijke relatie te zijn, maar omdat de beoordelingen van de annotaties ook aan andere factoren konden worden toegeschreven, kon niet worden vastgesteld dat de beoordelingen waren gerelateerd aan het constateren c.q. ophelderen van onduidelijke passages en motiveringsgebreken.³⁹ Op grond hiervan en in het licht van de bevindingen in de andere rechtsgebieden ligt het niet voor de hand aan te nemen dat dit kenmerk een indicator is voor de kwaliteit van annotaties.

Een vergelijkbare conclusie dient te worden getrokken voor het zoeken naar een verklaring voor de gewezen beslissing. In de bestuursrechtelijke en privaatrechtelijke annotaties leek het onwaarschijnlijk dat dit kenmerk gerelateerd is aan de kwaliteit van annotaties. Voor de strafrechtelijke annotaties werd geconcludeerd dat er eventueel wel een effect was, maar dat in de annotatie waaruit dit vermoeden is afgeleid, ook andere kenmerken in het spel waren die vermoedelijk ook of meer hebben bijgedragen aan de positieve waardering van de betreffende annotatie. De twijfels leiden dan ook tot de constatering dat er onvoldoende aanwijzingen zijn om te concluderen dat het zoeken naar een verklaring voor de gewezen uitspraak een goede voorspeller is voor de kwaliteit van annotaties.

Ook voor het vergelijken van de uitkomst met een uitkomst indien de rechter anders zou hebben besloten, geldt dat er te weinig aanwijzingen zijn om dit kenmerk als indicator te beschouwen voor de kwaliteit van annotaties. De omvang van de passages waarin dit aan de orde was, was gering. Bovendien

³⁹ De eventuele relatie tussen dit kenmerk en de kwaliteit van annotaties zou evenwel kunnen worden verklaard vanwege het gegeven dat één van de twee strafrechtelijke beoordelaars alle annotaties die op enigerlei wijze kritisch waren ten aanzien van de uitspraak als goed/uitstekend heeft aangemerkt.

bestond de indruk dat de positieve beoordelingen vermoedelijk eerder worden toegeschreven aan andere kenmerken dan aan de vergelijking.

3.3 *Kenmerken met ambigu beeld*

Het uitspreken van een toekomstverwachting ten aanzien van de ontwikkeling van een bepaald leerstuk bleek één keer wel te kunnen worden gerelateerd aan de kwaliteit van annotaties, en de anderen keren niet. Een verklaring hiervoor is dat de verwachting in eerstbedoelde annotatie gebaseerd was op, en onderbouwd werd aan de hand van, literatuur waarin kritiek op de huidige regel of toepassing is geuit, terwijl in de andere annotaties slechts een algemene verwachting werd uitgesproken die niet onmiddellijk voortvloeide uit de uitspraak (of uit het overige deel van de annotatie). Alleen in de eerste situatie lijkt dit kenmerk een voorspeller te zijn van een kwalitatief goede/uitstekende annotatie, maar niet in andere gevallen.

Het gemotiveerd aangeven of de uitspraak overtuigt, is een kenmerk dat over het algemeen niet, maar voor een enkele beoordelaar wel, gerelateerd lijkt te zijn aan de kwaliteit van annotaties. De analyse van bestuursrechtelijke en privaatrechtelijke annotaties leverde te weinig aanwijzingen op voor de conclusie dat het gemotiveerd aangeven of de uitspraak overtuigt van invloed is op de kwaliteit van annotaties. Bij de analyse van de strafrechtelijke annotaties bleek echter dat het kenmerk voor de ene beoordelaar wel, maar voor de andere beoordelaar geen criterium is geweest om annotaties als goed/uitstekend aan te merken. Deze observaties maken het gemotiveerd aangeven of de uitspraak overtuigt geen goede indicator voor de kwaliteit van annotaties.

4 **Resultaten fuzzy-set qualitative comparative analysis (fsQCA)**

Om te controleren of de vorenstaande indrukken en voorspellingen juist zijn, is een *fuzzy-set Qualitative Comparative Analysis* (fsQCA) uitgevoerd. Een fsQCA analyseert, net als veel andere onderzoekstechnieken, de relatie tussen uitkomstvariabelen (afhankelijke variabelen) en verklarende variabelen (onafhankelijke variabelen/predictoren), en brengt deze relatie tot uitdrukking in een cijfermatige schatting.⁴⁰ FsQCA is vooral geschikt voor situaties waarin de steekproef beperkt is.

FsQCA werkt zo dat de annotaties een score hebben op ieder van de kenmerken.⁴¹ Dit vereist een score op de verschillende kenmerken voor iedere annotatie tussen 0 (afwezig) en 1 (aanwezig). Hiervoor zijn de scores gebruikt die gebruikt zijn bij het opstellen van de matrices, waarbij een score van 0.5

⁴⁰ Rihoux & Ragin 2008.

⁴¹ Rihoux & Ragin 2008; Ragin 2000.

staat voor ‘kenmerk aangetroffen maar zeer beperkt in omvang in vergelijking met andere annotaties waar het kenmerk is aangetroffen’.⁴² De uitkomstvariabele – de beoordelingen zoals gegeven door de beoordelaars – is ook omgezet naar een score tussen 0 en 1, in die zin dat de score ‘1’ is toegekend aan annotaties die door beide/alle beoordelaars als goed/uitstekend zijn aangemerkt, de score ‘0’ aan annotaties die door geen beoordelaar als goed/uitstekend is aangemerkt en, bijvoorbeeld, de score ‘0.5’ aan annotaties die door één van de twee beoordelaars als goed/uitstekend is aangemerkt.⁴³

Eerst is bekeken of er bepaalde voorwaarden zijn wil een annotatie als kwalitatief goed/uitstekend worden aangemerkt (*necessity test*). Het beschrijven of samenvatten van de feiten en de casus enerzijds en het geven van een overzicht van het geldende recht anderzijds kwamen uit de analyse naar voren als twee kenmerken die annotaties hebben die als goed/uitstekend worden aangemerkt (*consistency* = 1.00, *coverage* = .52⁴⁴). Deze kenmerken zijn noodzakelijk, maar echter niet voldoende om van een kwalitatief goede/uitstekende noot te kunnen spreken. Daarom is tevens geanalyseerd welke kenmerken de kwaliteit van annotaties voldoende kunnen verklaren (*sufficiency tests*).

Meerdere modellen zijn getoetst. Het eerste model bestaat uit de kenmerken waarvan in de vorige fase van het onderzoek met redelijke mate van waarschijnlijkheid is vastgesteld dat zij gerelateerd zijn aan de beoordelingen van de deskundigen (model 1). Het gaat om de kenmerken schetsen van implicaties (*implic*), het nagaan welke leer van toepassing is c.q. wat geldend recht is (*leer*), het toetsen aan hoger recht (*hoger*), het ontwikkelen van een criterium, het vergelijken van dit criterium met het in de uitspraak gevolgde criterium (*crit*) en het systematisch nalopen van voor- en tegenargumenten (*syst*). Model 2 vult model 1 aan met twee kenmerken, te weten het (gemotiveerd) eens of oneens zijn met de uitspraak (*arg*) en het schetsen van toekomstverwachtingen (*toek*). Omdat sommige kenmerken alleen in bepaalde rechtsgebieden zijn aangetroffen of omdat zij soms gerelateerd leken aan een bepaald rechtsgebied, is dit meegenomen in de analyse.

Tabel 7 geeft de resultaten. Het is bij de interpretatie ervan van belang te bedenken dat de resultaten zijn gebaseerd op een relatief kleine steekproef en waarin bovendien alleen is gelet op aan-/afwezigheid van het kenmerk en op de lengte, wat hier als indicatie (proxy) voor diepgang en grondigheid dient. Daarnaast is het belangrijk te onderkennen dat het toetsen van het model en het interpreteren van de resultaten alleen zinvol is als daar een inhoudelijke basis voor is;

⁴² Zie Bijlage 4 voor de scores.

⁴³ Zie Bijlage 4 voor de scores.

⁴⁴ Beide kenmerken gaven dezelfde waarden voor *consistency* en *coverage*.

alleen of primair afgaan op getallen en coëfficiënten verdient hier (en overigens in het algemeen) geen aanbeveling.⁴⁵

Tabel 7 – Resultaten fsQCA

Model (kenmerken)	Coverage	Consistency
Model 1 (leer*sr*~pr*~br + hoger*br*~sr*~pr + implic*sr*~br*~pr + implic*crit*pr*~sr*~br + implic*br*~pr*~sr)	.51	.96
Model 2 (leer*sr*~pr*~br + hoger*br*~sr*~pr + implic*sr*~br*~pr + implic*crit*pr*~sr*~br + arg*sr*~br*~pr.law + toek*br*~sr*~pr)	.61	.91

Noot: Intermediate solutions worden weergegeven. * = combinatie. ~ = afwezigheid. Br = bestuursrecht. Sr = Strafrecht. Pr = Privaatrecht. Frequency cutoff: 1.00. Consistency cutoff: 0.80.

Box 3 – Qualitative Comparative Analysis (QCA)

Qualitative Comparative Analysis is een techniek bedoeld voor het oplossen van problemen die worden veroorzaakt doordat causale uitspraken worden gedaan op basis van een klein aantal gevallen of observaties.

Model 1 heeft een *consistency value* van .96. De waarde van model 2 is .91. Over het algemeen worden waarden tussen .80 en .99 als acceptabel gezien.⁴⁶ Beide zijn dus betrouwbaar te noemen. Ook de gevonden *coverages* van .51 (model 1) en .61 (model 2) zijn behoorlijk te noemen, zeker gezien de complexiteit van de materie (kwaliteit van onderzoek) en de omstandigheid dat alleen naar aan-/afwezigheid en omvang van de passages is gekeken.⁴⁷ Op grond van de coëfficiënten zou model 1 te prefereren zijn vanwege de hogere consistency value en model 2 voor haar hogere coverage, maar als gezegd dienen de uitkomsten ook inhoudelijk te kunnen worden onderbouwd.

De resultaten van de modellen, die overigens in belangrijke mate overeenkomen, sluiten aan bij eerdere constatering. Het toetsen aan hoger recht (bestuursrecht) en het nagaan welke leer is toegepast (strafrecht) hebben in beide modellen alsook in de eerdere analyse een effect op de uitkomst. Voor het schetsen van toekomstverwachtingen werd voor bestuursrechtelijke annotaties ver-

⁴⁵ Ragin 2000, p. 246.

⁴⁶ Rihoux & Ragin 2008; Ragin 2000. *Consistency values* zijn vergelijkbaar met p-waarden in statistische analyses, die een schatting geven van de kans dat de resultaten op toeval berusten, zie Ragin 2006.

⁴⁷ *Coverage* is vergelijkbaar met *R-squared* waarden in regressieanalyse, die een weergeven hoe goed de uitkomsten (scores) kunnen worden verklaard met de predictoren (kenmerken), zie Ragin 2006.

moed dat het de beoordelingen deels kon verklaren. De fsQCA bevestigt dat beeld.

Verder laat model 2 zien dat het (gemotiveerd) eens of oneens zijn met de uitspraak een positieve invloed heeft op de kwaliteit van de annotatie. Hiervan werd al eerder vastgesteld dat dit vermoedelijk kan worden verklaard doordat één van de twee beoordelaars alle annotaties die kritisch waren ten opzichte van de uitspraak als goed/uitstekend heeft aangemerkt.

Voor het schetsen van implicaties laat model 1 een effect zien in het strafrecht, het bestuursrecht en het privaatrecht, maar bij privaatrechtelijke annotaties alleen in combinatie met het ontwikkelen en toetsen van een criterium.⁴⁸ Model 2 laat geen effect zien voor de bestuursrechtelijke annotaties. Vanwege eerder genoemde beperkingen is het de vraag of veel aandacht moet worden geschonken aan dergelijke verschillen. Het algemene beeld is dat zowel model 1 als model 2 een goed verklaringsmodel vormt en, belangrijker, in belangrijke mate overeenstemt met wat eerder in deze studie is opgevallen en geconcludeerd.

5 Conclusie

De analyse die in dit hoofdstuk is uitgevoerd, laat zien dat er wel degelijk regelmatigigheden schuilgaan achter annotaties. Dit bevestigt het vermoeden dat het moeilijk is voor deskundigen om adequaat te formuleren wat een annotatie tot een goede annotatie maakt.

De resultaten laten een patroon zien waarbij bepaalde annotaties een grotere kans hebben om als kwalitatief goed of uitstekend te worden aangemerkt dan andere annotaties. Het gaat om annotaties waarbij sprake is van, wat ik hier noem, normatieve diepgang. Over het algemeen gaat het om constatering, conclusies en aanbevelingen voor de interpretatie, toepassing en ontwikkeling van het recht.

Wat valt, meer concreet, onder normatieve diepgang? Ik tracht de grenzen van dit criterium af te tasten aan de hand van de gevonden indicatoren. Onder normatieve diepgang valt het analyseren van implicaties van de uitspraak, het toetsen aan ‘hoger’ recht, het nagaan welke leer is toegepast, het ontwikkelen van een criterium (en toetsen of de rechter dit criterium heeft toegepast) en het systematisch (uitputtend) bespreken van voor- en tegenargumenten, bij voorkeur over een onderwerp dat van betekenis is in het juridische domein.

De indicatoren hebben normatieve betekenis. Indien de annotator bijvoorbeeld een criterium opstelt, toetst en behoorlijk onderbouwt, is het bruikbaar om in andere, vergelijkbare gevallen toe te passen. Op een vergelijkbare

⁴⁸ Een vergelijking tussen de *intermediate solution* en de *parsimonious solution* laat bovendien zien dat de combinatie crit*implic geen grotere *coverage* of *consistency* oplevert dan een oplossing met alleen crit.

wijze kan het toetsen aan ‘hoger’ recht ertoe leiden dat beslissingen in de toekomst anders dienen te zijn of dat expliciet(er) rekening wordt gehouden met het betreffende beginsel of verdrag, en kan het schetsen van implicaties van een uitspraak nopen tot nuancering of aanpassing van de gewezen regel. Niet onder het criterium van normatieve diepgang valt het toetsen of ter discussie stellen van opvattingen van andere auteurs, tenzij het gaat om concurrerende theorieën waarover een uitspraak duidelijkheid schept, in die zin dat de uitspraak aangeeft welke theorie een grotere verklarings- en voorspellingskracht heeft. In het laatste geval heeft een conclusie betekenis voor de interpretatie en toepassing van het recht.

Het criterium van normatieve diepgang brengt verder mee dat, wil een annotatie als kwalitatief hoogstaand kunnen worden aangemerkt, de annotatie het individuele geval – de besproken uitspraak – ontstijgt. Dit inzicht verklaart vermoedelijk waarom annotaties over het algemeen niet als kwalitatief goed of hoogstaand worden aangemerkt als zij voor het overgrote deel bestaan uit het beschrijven of samenvatten van de feiten of de uitspraak, uit het schetsen van kwesties die zich geregeld voordoen, of uit het geven van een overzicht van het geldende recht en de literatuur daarover. Ook verklaart het waarom het gemotiveerd eens of oneens zijn met de uitspraak (anders dan door het schetsen van implicaties, toetsen aan ‘hoger’ recht, etc.) meestal niet tot een hoge waardering van een annotatie leidt. Bij dit kenmerk gaat het vaak om een kritische evaluatie van de uitspraak zelf, zonder dat conclusies worden getrokken die in breder verband relevant kunnen zijn. Hetzelfde geldt voor het ophelderen van onduidelijke passages (tenzij de implicaties van deze onduidelijkheden grondig worden besproken), het opmerken van motiveringsgebreken (*idem*) en voor het opwerpen van vragen, vaak zonder deze te beantwoorden of adequaat te beantwoorden.

Het voorgaande betekent uiteraard niet dat uitspraken met indicatoren die positief met de kwaliteit van annotaties zijn geassocieerd, zonder meer tot een hogere kwaliteit van annotaties leidt. De analyse dient tevens een behoorlijke diepgang te hebben. Ook al is het begrip ‘diepgang’ niet precies te omschrijven – een precies oordeel vereist in de regel een oordeel van een deskundige in het vakgebied – geven de uitgevoerde analyses hier enige richtlijnen. Zo is onvoldoende het vergelijken met eerdere uitspraken door louter overeenkomsten of verschillen te constateren (‘de uitspraak sluit wel/niet aan bij eerdere uitspraken’). De vergelijking dient verder te gaan, bijvoorbeeld door de gevolgen van de uitspraak door te rekenen of door deze in verband te brengen met andere rechtsgebieden of rechtsstelsels (verdragen, interne of externe rechtsvergelijking, etc.).

Uit het vorige hoofdstuk leek reeds te volgen dat diepgang verder meebrengt dat bij een annotatie, wil deze als goed/uitstekend worden aangemerkt,

sprake moet zijn van zorgvuldig bronnengebruik (voldoende en juiste bronnen) en dat de uitspraak alsook eventuele achtergronden adequaat worden beschreven. Daarnaast moeten conclusies die tot doel hebben de gevolgen door te trekken naar andere leerstukken of rechtsgebieden, ook redelijkerwijs door te trekken zijn. In een van de geanalyseerde annotaties bleken hier twijfels over te bestaan, met als gevolg dat de waardering lager uitviel in vergelijking met andere annotaties waarin conclusies zijn getrokken die verder reikten dan de uitspraak in kwestie.

In het kader van het element van diepgang is nog het volgende van belang. In de bestudeerde annotaties zijn de passages met kenmerken die positief geassocieerd zijn de kwaliteit van annotaties van een behoorlijke (niet-geringe) omvang. Hoogstwaarschijnlijk gaat het hier niet zozeer om de omvang, maar staat de omvang voor een zekere grondigheid en diepgang. Hoewel het bronnengebruik enig houvast biedt, zijn deze elementen moeilijk op een objectieve wijze te meten. De beoordeling zal voor een belangrijk deel afhangen van de kennis, ervaring en opvatting van de deskundigen in het betreffende rechtsgebied of met betrekking tot het betreffende leerstuk. Een langere noot zal niet zonder meer leiden tot een hogere kwaliteit. Sterker, het is niet ondenkbaar dat het vergroten van het aantal woorden een tegenovergesteld effect (lagere kwaliteit) kan hebben.

De analyse die in dit hoofdstuk is uitgevoerd, doet sterk vermoeden dat normatieve diepgang zoals hier omschreven een belangrijk criterium is om de kwaliteit van annotaties te beoordelen. Dit neemt niet weg dat als het schort aan andere elementen, zoals structuur en leesbaarheid, dit ertoe kan leiden dat de annotatie als kwalitatief minder hoogstaand wordt aangemerkt. Bovendien geldt dat het criterium louter een richtsnoer is. Er zullen altijd annotaties zijn waarin niet is voldaan aan dit criterium en die toch als kwalitatief hoogstaand worden aangemerkt. Deze uitzonderingen zijn echter moeilijk te duiden. Zo kon niet worden achterhaald waarom in bepaalde annotaties het samenvatten van de uitspraak en de literatuur daarover positief geassocieerd was met de kwaliteit van annotaties en in andere gevallen niet.

Experimentele studies naar de kwaliteit van annotaties

1 Inleiding

De analyses zoals uitgevoerd in de voorgaande hoofdstukken laten zien dat er wel degelijk indicatoren zijn aan de hand waarvan de kwaliteit van annotaties kan worden vastgesteld, maar dat deze indicatoren afwijken van en specifiek zijn dan de criteria die worden genoemd binnen het juridisch-academisch forum. Het relateren van kwaliteitsoordelen over annotaties aan kenmerken van annotaties zoals die in de literatuur zijn gedefinieerd, bleek immers niet de verschillen te verklaren tussen goede/uitstekende en matige/slechte annotaties.

Voor sommige van de gevonden indicatoren bleken er sterke aanwijzingen te zijn dat zij geassocieerd zijn met de kwaliteit van annotaties. Dit gold bijvoorbeeld voor het schetsen van implicaties en het toetsen aan ‘hoger’ recht (bestuursrecht).

Een belangrijke beperking van de in het vorige hoofdstuk uitgevoerde analyse is dat het niet zeker is dat de gevonden effecten ook daadwerkelijk causaal zijn. De gevonden indicatoren lijken weliswaar gerelateerd (gecorrleerd) aan de kwaliteit van annotaties, maar het bewijs voor causale effecten is beperkt. Niet kan worden uitgesloten dat andere kenmerken de kwaliteit van de onderzochte annotaties kunnen verklaren, of dat er (andere) storende factoren in het spel waren.

Om meer zicht te krijgen op of de gevonden invloed van de gevonden indicatoren op de kwaliteit van annotaties causaal is, is een tweetal experimentele studies uitgevoerd. Vanwege een verwacht laag aantal respondenten zijn twee van de gevonden indicatoren getoetst.

2 Studie 1 – Schetsen van implicaties en de kwaliteit van annotaties

2.1 Onderzoeksvraag en opzet

Uit de analyse zoals uitgevoerd in het vorige hoofdstuk bleek dat het schetsen van implicaties een indicatie blijkt te zijn voor een kwalitatief goede annotatie, mits sprake is van voldoende diepgang en grondigheid, alsook van een juiste analyse. In een aantal gevallen bleek echter ook dat een weergave van de feiten, de uitkomst en/of het geldende recht dat op het geval van toepassing is soms kan verklaren waarom een annotatie als goed/uitstekend is aangemerkt. Omdat in (vrijwel) alle van de onderzochte annotaties sprake is van het weergeven van

feiten, de uitkomst en/of van het geldende recht, kon niet worden uitgesloten dat het schetsen van implicaties een positief effect heeft op de kwaliteit van annotaties.

Om dit na te gaan, is een experiment opgezet waarbij de deelnemers aan het experiment in twee groepen zijn verdeeld en een annotatie is voorgelegd. In de ene groep bevatte de annotatie passages waarin de annotator de implicaties van de uitspraak bespreekt. In de andere groep ontbraken deze passages. Voor het overige waren de annotaties identiek. Door de deelnemers vervolgens de kwaliteit te laten scoren en de scores van de twee groepen te vergelijken, kon worden vastgesteld of de annotatie zonder het schetsen van implicaties lager zou scoren dan de annotatie met het schetsen van implicaties, en dus of dit kenmerk een (positief) effect heeft op de kwaliteit van annotaties, en zo ja, hoe groot dit effect is. Om te onderzoeken of een eventueel effect het gevolg is van andere factoren, zoals een verschil in het aantal woorden – bij het weglaten van passages wordt een annotatie minder omvangrijk – of een gebrekkige structuur als gevolg van het weglaten van passages, zijn dergelijke aspecten gemeten en dus meegenomen in het experiment.

Uit de analyse zoals uitgevoerd in het vorige hoofdstuk volgt het vermoeden dat het schetsen van implicaties leidt tot een hogere kwaliteit van de annotatie. De hypothese die dan ook is getoetst, is:

H₁: Het schetsen van implicaties van de uitspraak in een annotatie leidt tot een betere beoordeling van de annotatie.

2.2 Methode

Sample

De annotatie met bijbehorende vragenlijst is in elektronische vorm verstuurd aan 197 personen. De deelnemers waren gepromoveerde juristen verbonden aan de vakgroep privaatrecht⁴⁹ van de Nederlandse rechtenfaculteiten. De verwachting was dat gepromoveerden in een bepaald rechtsgebied voldoende in staat, en daarmee geschikt, zouden moeten zijn om de kwaliteit van een annotatie te beoordelen. Gepromoveerden die betrokken waren in een eerdere fase van het onderzoek zijn niet benaderd.

Van de 197 respondenten die zijn benaderd, beantwoordde 20% (N=40) een deel van de vragenlijst. Van de aangeschreven respondenten vulde 14% (N=28) de volledige vragenlijst in. Twee respondenten braken de vragenlijst af bij de tweede vraag, namelijk bij het geven van een korte samenvatting van de annotatie. Acht respondenten haakten vervolgens af nadat zij de annotatie hadden beoordeeld. Daarna brak één respondent de vragenlijst af toen gevraagd

⁴⁹ Waaronder tevens begrepen burgerlijk recht en civiel recht.

werd naar algemene opvattingen over annotaties. Aan het slot van de vragenlijst werd gevraagd naar persoonskenmerken. Eén respondent brak de vragenlijst af toen werd gevraagd naar leeftijd. Deze respondent is niet meegenomen in deze analyse. Uit verkennende analyses bleek dat leeftijd een rol speelt respectievelijk kan spelen bij de beoordeling van annotaties.

Daarnaast bleek één respondent niet gepromoveerd te zijn. Deze respondent is daarom buiten de analyse gelaten. Verder leverde een inspectie van de data op dat één van de respondenten als een uitschieter ('outlier') te beschouwen is. De beoordeling van deze respondent ligt meer dan drie standaarddeviaties onder het gemiddelde en drie 'interkwartielen' onder het eerste kwartiel.

Box 4 – Uitschieters (outliers)

Uitschieters zijn waarnemingen die niet lijken te passen bij de andere waarden. Zij kunnen patronen verstoren of in belangrijke mate veranderen en daarmee de resultaten van analyses wezenlijk beïnvloeden.

Een manier om uitschieters te duiden, is om de gemiddelde afwijking van het gemiddelde (standaarddeviatie) te berekenen en vervolgens na te gaan hoe veel standaarddeviaties de waarneming afwijkt van het gemiddelde. Bij een afwijking van meer dan drie standaarddeviaties wordt over het algemeen van een uitschieter gesproken.

Een andere manier is om de waarnemingen in vier gelijke delen op te delen en te berekenen hoe ver de uitschieter verwijderd ligt van het eerste kwartiel (bij extreem lage scores) of van het derde kwartiel (bij extreem hoge scores). Indien de waarneming meer dan 1,5 'interkwartiel' (het verschil tussen het eerste kwartiel en het derde kwartiel) verwijderd is van het eerste respectievelijk derde kwartiel, wordt over het algemeen van een uitschieter gesproken.

Meer in het algemeen bleek deze respondent er een interessante opvatting op na te houden, zo bleek uit de toelichting die hij/zij gaf:

'Rechtspraak gaat meestal over vragen die vrij simpel in gewone woorden te vertellen zijn. Als de rechter dat zelf zou doen, plus aan zou geven hoe de uitspraak in de lijn past van eerdere uitspraken, zou er geen enkele behoefte aan annotaties zijn. Behoeft aan annotaties is ook sterk afgenomen omdat uitspraken eenvoudig op allerlei manieren opzoekbaar zijn. Het is dus helemaal niet meer nodig dat iemand de eerdere relevante uitspraken erbij zoekt. Dat doet de computer wel. Het is dus ook een genre dat ontbreekt in landen waar uitspraken in een goede database zitten en waar rechters een beetje normale mensentaal spreken en to the point weten te komen. De aanwezigheid van annotaties zegt iets over de kwaliteit van rechterlijke uitspraken. Achter de annotatie zit een circuit van uitgevers die tijdschriften willen vullen en wetenschappers die daar ontmoetingsplekken hebben (redacties, etentjes, etc.). Het zou interessant zijn dat te ontrafelen.'

De gegevens van deze respondent zijn dan ook niet in de analyses betrokken.⁵⁰ Dit brengt het aantal respondenten waarmee de analyses zijn uitgevoerd op 26. De respondenten gaven allen een adequate samenvatting van de annotatie.

Van deze 26 respondenten was 69% (N=18) man en 31% (N=8) vrouw. De gemiddelde leeftijd was 47 jaar. Van de respondenten gaven er 9 aan voltijd hoogleraar te zijn (waarvan één 'bijna voltijd'), 4 personen bijzonder hoogleraar, 7 respondenten universitair hoofddocent en 6 personen universitair docent. Verder vulden 7 van de 26 respondenten in dat zij vóór 2000 gepromoveerd waren, terwijl 15 respondenten hebben aangegeven in 2000 of daarna te zijn gepromoveerd. Van drie respondenten is het promotiejaar onbekend. Eén respondent gaf 1900 als promotiejaar op. Allen bevestigden gespecialiseerd te zijn in het privaatrecht.

De annotatie die als basis diende voor dit experiment, is annotatie nr. 2 (privaatrecht). In deze annotatie beslaat het schetsen van implicaties een behoorlijk deel van de annotatie (ruim 20%). Bovendien bleek het voor deze annotatie relatief eenvoudig om de passages weg te laten uit de annotatie zonder de structuur ervan aan te tasten.

Procedure

De deelnemers dienden de annotatie te lezen en vervolgens de vragenlijst in te vullen, waarin ook is getoetst of de manipulatie – het opnemen of weglaten van bepaalde passages – is geslaagd. Een deel van de deelnemers kreeg een annotatie voorgelegd met de passages waarin het schetsen van implicaties aan de orde was. Bij de andere groep ontbraken de passages waarin dit aan de orde was. Voor het overige waren de voorgelegde annotaties identiek. De deelnemers zijn random over de groepen verdeeld. Dit heeft ertoe geleid dat de groep die de volledige annotatie te lezen kreeg, bestond uit 11 personen, en de groep die de onvolledige annotatie heeft gelezen uit 15 personen. Er werden geen grote verschillen gevonden tussen de groepen in termen van leeftijd, geslacht, type aanstelling (hoogleraar, UHD, UD) en gemiddeld promotiejaar.

Na het lezen van de annotatie dienden de deelnemers een kwaliteitsoordeel te geven over de betreffende annotatie. Zij dienden dit op een Likertschaal (zevenpuntschaal) aan te geven. Vervolgens werden de deelnemers verzocht om een hoeveelheid stellingen te beoordelen. Hiermee kon worden vastgesteld waarop de respondenten hun oordeel hadden gebaseerd en of eventuele verschillen in scores te relateren zouden zijn aan het schetsen van implicaties. De stel-

⁵⁰ Het is niet mogelijk om de analyse goed uit te voeren met alle respondenten. Nog afgezien van de uitschieter, zal hierna blijken dat het in de analyse belangrijk is om rekening te houden met andere factoren dan alleen het schetsen van implicaties. Het heeft daarom weinig zin om de analyse uit te voeren met alle respondenten indien zij voor die andere factoren geen scores hebben.

lingen hadden betrekking op de leesbaarheid/structuur/omvang, het bronnengebruik, de relevantie van de annotatie, de beschrijving van de uitspraak en de daarin gewezen rechtsregel. Ook zijn stellingen opgenomen over het schetsen van implicaties, de opvatting over annotaties, de affiniteit met het vakgebied, de affiniteit met het schrijven van annotaties en stellingen over de manipulatie met de versie (onvolledig/volledig).

2.3 Resultaten

Van de 26 respondenten hebben er 11 de volledige annotatie beoordeeld, dat wil zeggen de annotatie met daarin passages over het schetsen van implicaties. De andere 15 respondenten hebben de onvolledige annotatie gelezen, dit wil zeggen de annotatie waarin de passages zijn weggelaten waarin implicaties worden geschetst. De resultaten laten, zoals verwacht, zien dat de annotatie waarin de passages zijn weggelaten waarin de implicaties van de uitspraak worden besproken, minder goed wordt beoordeeld [$M = 5.33$, $SD = .72$] dan de annotatie met de passages [$M = 5.73$, $SD = .79$]. Een verschil van 0.40 op een zevenpuntschaal lijkt klein, maar een inspectie van de data leert dat de laagste score 4 is en de hoogste score 7. Het gevonden verschil bedraagt daarmee ruim 13% van het verschil tussen de hoogste en laagste score.

De hoge scores ten aanzien van de beoordelingen kunnen worden verklaard doordat een annotatie is geselecteerd die als kwalitatief hoogstaand is aangemerkt, zelfs als het schetsen van implicaties zou ontbreken. De resultaten kunnen echter ook suggereren dat de respondenten die hebben meegedaan aan het onderzoek, over het algemeen een positieve houding hebben ten aanzien van annotaties. Het laatste wordt ondersteund door additionele data. Aan de hand van vier stellingen is getracht te achterhalen wat in het algemeen de opvatting was van respondenten over annotaties. Deze stellingen hadden een redelijke tot goede betrouwbaarheid (Cronbach's Alpha = .72).⁵¹

Box 5 – Cronbach's Alpha

Cronbach's Alpha is een maat om de interne consistentie van een hoeveelheid items (vragen of stellingen) te meten. Indien meerdere items worden gebruikt om hetzelfde concept te meten, kan aan de hand van Cronbach's Alpha worden nagegaan of de items daadwerkelijk hetzelfde meten. In onderzoek beschouwen sommigen een score van .70 of hoger als acceptabel. Anderen houden een grens aan van .80.

De twee items die het meest direct de opvattingen van respondenten over annotaties in het algemeen maten, zijn redelijk hoog, te weten 4.81 en 5.23, gemeten op een zevenpuntschaal (Tabel 8). Mogelijk hebben vooral degenen die over het

⁵¹ Eén item is gespiegeld ten behoeve van de berekening.

algemeen positief tegenover annotaties staan meegewerkt aan dit onderzoek. Dit zou betekenen dat de resultaten naar die groep kunnen worden gegeneraliseerd, maar mogelijk niet naar gepromoveerden die een minder positieve (of zelfs negatieve) opvatting hebben ten aanzien van annotaties.

Tabel 8 – Scores op items opvattingen over annotaties

	M	SD
Gemiddeld genomen beoordeel ik de kwaliteit van annotaties als ... (1=zwak, 7=uitstekend)	4.81	.98
Annotaties dienen bij onderzoeksbeoordelingen als academische publicatie te worden aangemerkt (1=helemaal mee oneens, 7=helemaal mee eens)	5.23	1.77

Noot: De andere twee items hadden betrekking op de kwaliteit van annotaties in vergelijking met dissertaties resp. op de stelling dat de kwaliteit van annotaties sterk verschilt tussen annotaties onderling.

Een ander mogelijk probleem was dat respondenten de annotatie die zij moesten beoordelen, al eerder hadden gelezen. Dit zou een effect kunnen hebben op hoe de respondenten de annotatie zouden scoren (waarderen). Om hier rekening mee te kunnen houden, is aan de respondenten gevraagd of zij wisten wie de annotator was van de noot die zij hadden gelezen, en zo ja, of zij de naam van de annotator konden geven. Van de respondenten bleek 39% (N=10) te weten wie de annotator was en noemde daar ook de juiste naam bij. Hiervan hadden 3 respondenten de onvolledige versie gelezen en 7 respondenten de volledige versie.

Om te toetsen of het gevonden verschil in beoordeling van .40 ten aanzien van de beoordeling van de versies van de annotaties niet op toeval berust, is een regressieanalyse uitgevoerd.

Box 6 – Regressieanalyse

Regressieanalyse is een statistische techniek om na te gaan of er een relatie bestaat tussen twee of meer variabelen. Hiermee kan bijvoorbeeld worden gemeten of er statistisch gezien een samenhang is tussen de beoordeling van annotaties en bijvoorbeeld de deskundigheid van de beoordelaar enerzijds en de versie die deze heeft gelezen anderzijds.

Getoetst is of het voor de beoordeling van de annotatie uitmaakt welke versie de respondenten hadden gelezen. Vanwege de hiervoor genoemde reden is daarbij rekening gehouden met de omstandigheid of de respondent wist wie de annotator was van de annotatie die ter beoordeling is voorgelegd.

Daarnaast zou de beoordeling door de respondenten mogelijk beïnvloed kunnen zijn door hun algemene opvatting over annotaties. Daarom is ook dit element meegenomen, als gezegd in de vorm van een viertal stellingen (Cronbach's Alpha = .72).

Voorts was de verwachting dat de affiniteit van de beoordelaar met het vakgebied van invloed zou kunnen zijn op de beoordeling. De affiniteit met het vakgebied is gemeten met een zevental items. Deze items gingen over in hoeverre de respondenten het geldende recht (literatuur en/of jurisprudentie) bijhielden over het vakgebied en over het onderwerp in kwestie, in hoeverre zij gedoceerd hebben over het vakgebied en het onderwerp in kwestie, hoe zij hun deskundigheid beoordeelden ten aanzien van de problematiek die in de annotatie aan de orde is en over hoe veel publicaties van de respondenten betrekking hadden op de thematiek uit de uitspraak/annotatie. De betrouwbaarheid van de items hiervan bleek hoog (Cronbach's Alpha = .92), wat suggereert dat de items op een betrouwbare wijze het concept van affiniteit met het vakgebied meten. Verder leek ook de affiniteit met het schrijven van annotaties relevant. Ook hierover is een item opgenomen.

De items van de genoemde concepten (bekendheid met annotator, opvattingen, affiniteit vakgebied, affiniteit schrijven annotaties) zijn, in aanvulling op het meten van de betrouwbaarheid met Cronbach's Alpha, onderworpen aan een *Principale Componentenanalyse* (PCA). Een PCA maakt het mogelijk om een hoeveelheid items terug te brengen tot een aantal, meer algemene factoren. Uit de uitgevoerde PCA blijkt dat bepaalde items inderdaad de opvattingen over annotaties en affiniteit met het vakgebied meten (Tabel 9). Op basis van de SCREE-plot zijn twee algemene componenten geselecteerd, die samen 59,2% van de variantie verklaren. De componentladingen voor de items die de twee factoren vertegenwoordigen, blijken (voldoende) hoog (gelijk aan of groter dan .65).

Tabel 9 – Componentladingen PCA

Item	Affiniteit vakgebied (component 1)	Opvatting annotaties (component 2)
In hoeverre houdt u het geldende recht (literatuur en/of jurisprudentie) bij m.b.t. [rechtsgebied]? (1=helemaal niet, 7=helemaal wel)	.88	-.27
In hoeverre houdt u het geldende recht (literatuur en/of jurisprudentie) bij m.b.t. [onderwerp]? (1=helemaal niet, 7=helemaal wel)	.90	-.08
In hoeverre heeft u gedoceerd over [rechtsgebied]? (1=helemaal niet, 7=heel vaak)	.85	-.09
In hoeverre heeft u gedoceerd over [onderwerp]? (1=helemaal niet, 7=heel vaak)	.79	.07
Hoe beoordeelt u uw deskundigheid ten aanzien van de problematiek die in de annotatie aan de orde is? (1=helemaal niet deskundig, 7=heel deskundig)	.90	-.10
Hoe veel van uw publicaties gaan over [rechtsgebied]? (0= geen, 1= 1-5, 2= 6-10, 3= >10)	.83	-.30
Hoe veel van uw publicaties gaan over [onderwerp]? (0= geen, 1= 1-5, 2= 6-10, 3= >10)	.77	.19
Hoe veel annotaties schrijft u jaarlijks (bij benadering)? (Het gaat om alle annotaties die u schrijft, dus ook om annotaties die over andere onderwerpen gaan dan [...]) (0= geen, 1= 1-5, 2= 6-10, 3= >10)	.19	.05
Gemiddeld genomen beoordeel ik de kwaliteit van annotaties als ... (1=zwak, 7=uitstekend)	.09	.69
Annotaties dienen bij onderzoeksbeoordelingen als academische publicatie te worden aangemerkt (1=helemaal mee oneens, 7=helemaal mee eens)	.32	.74
De kwaliteit van annotaties verschilt sterk tussen annotaties onderling (1=helemaal mee oneens, 7=helemaal mee eens) (gespiegeld)	-.08	.68
Annotaties zijn even relevant als dissertaties (1=helemaal mee oneens, 7=helemaal mee eens)	.23	.77

Noot: Ladingen gelijk aan of groter dan .50 staan vetgedrukt. Rechtsgebied en onderwerp zijn niet weergegeven met het oog op het kunnen achterhalen van de persoon van de annotator.

Box 7 – Principale Componentanalyse (PCA)

Een PCA heeft in essentie tot doel om de oorspronkelijke items (vragen/stellingen) terug te brengen tot een kleiner aantal componenten. De componenten bestaan dan ook uit variabelen die hoog laden op de betreffende component en (bij voorkeur) laag laden op de andere component(en). In deze studie worden ladingen gelijk of groter dan .50 gehanteerd als ondergrens.

Waar verder rekening mee dient te worden gehouden, is dat de onvolledige noot anders leest dan de volledige noot en dat dit verschil tot uitdrukking kan komen in de scores, bijvoorbeeld omdat de structuur van de onvolledige annotatie niet deugt, omdat de omvang te klein/groot is. Het gevonden verschil zou in dat geval niet (zozeer) toe te schrijven zijn aan de aan- dan wel afwezigheid van het element van het schetsen van implicaties.

Aan de hand van een vijftal stellingen is getracht te achterhalen wat de invloed van dergelijke structuurelementen is geweest op de beoordelingen van de respondenten. De stellingen, die een hoge betrouwbaarheid opleverden (Cronbach's Alpha = .85), gingen over de vraag in hoeverre de respondenten hun beslissing hadden gebaseerd op het feit dat er iets mankeerde aan de omvang, structuur, leesbaarheid en/of het taalgebruik van de annotatie.

Vervolgens is een regressieanalyse uitgevoerd met de beoordeling van de annotaties als afhankelijke variabele en de versie (onvolledig / volledig) als afhankelijke variabele (model 1). Vervolgens is ook een analyse uitgevoerd met de variabelen van model 1, aangevuld met de predictoren bekendheid met de annotator, affiniteit met het vakgebied, affiniteit met het schrijven van annotaties, opvatting over annotaties en leesbaarheid/structuur/omvang en leeftijd (model 2). Tabel 10 geeft de resultaten ten aanzien van de modellen.

Tabel 10 – Resultaten regressieanalyse

Variabele	Model 1		Model 2	
	B (s.e.)	β	B (s.e.)	β
Constant	4.939 (.449)	-	5.539 (.782)	-
Versie (onvolledig/volledig)	.394 (.298)	.26	.540 (.246)*	.36
Bekendheid met annotator (nee/ja)			.215 (.297)	.14
Affiniteit vakgebied			-.082 (.099)	-.17
Affiniteit schrijven annotaties			.519 (.217)*	.40
Opvatting over annotaties			.027 (.109)	.04
Leesbaarheid/structuur/omvang (perceptie)			-.165 (.102)	-.33
Leeftijd			-.017 (.011)	-.26
R²	.07		.54	
F	1.74		2.99	
N	26		26	

Noot: * $p < .05$.

De resultaten laten zien dat de versie (onvolledig/volledig) een effect heeft op de beoordelingen van de annotatie door de respondenten indien rekening wordt gehouden met de hiervoor genoemde aspecten, zoals affiniteit met het vakgebied. Het effect is niet alleen statistisch significant, wat wil zeggen dat de kans 5% of kleiner is dat het gevonden effect op toeval berust ($p = .04$), maar is tevens groot te noemen ($\beta = .357$). Hetzelfde geldt voor affiniteit met het schrijven van annotaties ($p = .03$, $\beta = .401$), in die zin dat naarmate beoordelaars zelf vaker annotaties schrijven, zij een annotatie positiever zullen beoordelen.

De resultaten geven geen indicatie van een effect van de andere predictoren, maar dat wil niet zeggen dat er geen effect zou zijn. Waarschijnlijk zouden bij een groter aantal respondenten meer effecten zijn gevonden. Met name valt een effect te verwachten voor leeftijd ($p = .13$, $\beta = -.263$), in die zin dat de waardering voor een annotatie lager uitvalt naarmate de beoordelaar een hogere leeftijd heeft. Met andere woorden, indien de gevonden trend in de data zich zou voortzetten bij een groter aantal respondenten, is de verwachting dat ook leeftijd een effect heeft op de beoordeling van annotaties. Verder is het denkbaar dat hoe beoordelaars annotaties in het algemeen waarderen een effect heeft op de beoordeling van een annotatie. Of dat zo is, kon niet worden getoetst. De respondenten die hebben meegedaan aan deze studie waren positief over de kwaliteit van annotaties in het algemeen. Dit verklaart waarom geen effect van de predictor 'opvatting over annotaties' is gevonden. Tot slot zou er bij een hogere respons vermoedelijk een effect zijn van leesbaarheid/structuur/omvang van de annotatie, maar dit heeft hier minder inhoudelijke betekenis, omdat deze predictor is opgenomen om het effect van de versie (onvolledig/volledig) te onderscheiden van het effect van leesbaarheid/structuur/omvang.

Aan de hand van stellingen is getracht de redenen te achterhalen waarop de respondenten hun beoordeling hadden gebaseerd. De stellingen hadden betrekking op de concepten leesbaarheid/structuur/omvang (5 items, Cronbach's Alpha = .85), beschrijving (12 items, Cronbach's Alpha = .86), relevantie (reikwijdte/nut/nieuwswaarde) (7 items, Cronbach's Alpha = .83), bronnengebruik (5 items, Cronbach's Alpha = .84) en implicaties van de uitspraak (12 items, Cronbach's Alpha = .96). Ondanks de hoge betrouwbaarheidsscores, laat een additionele Principale Componentanalyse zien dat de concepten moeilijk te onderscheiden zijn. Er is nauwelijks overeenstemming tussen de componenten uit de PCA en de zojuist genoemde concepten.⁵² Dit is in lijn met het beeld uit de vorige hoofdstukken, namelijk dat beoordelaars niet goed in staat zijn om de redenen aan te geven waarop zij hun beslissing hebben gebaseerd. Het toevoegen van deze concepten als predictor levert dan ook geen verbetering op van het

⁵² Om deze reden zijn de resultaten niet opgenomen.

verklaringsmodel.⁵³ Wel leidt het opnemen van alleen de predictor ‘implicaties’ (ten opzichte van model 2) ertoe dat het effect van versie (onvolledig/volledig) niet-significant wordt. Dit ondersteunt de gedachte dat het schetsen van implicaties inderdaad een effect heeft op de beoordeling van annotaties.

2.4 Discussie

Het schetsen van implicaties van de uitspraak in een annotatie blijkt, zoals eerder in deze studie werd vermoed, te leiden tot een betere beoordeling van een annotatie (H_1), onder constanthouding van affiniteit met het vakgebied, affiniteit met het schrijven van annotaties, opvatting over annotaties, bekendheid met de annotator, leeftijd en leesbaarheid/structuur/omvang. Ondanks het lage aantal respondenten, is het effect van het schetsen van implicaties zowel statistisch significant als sterk. Ook affiniteit met het schrijven van annotaties heeft een positief effect op de beoordeling van een annotatie. Naarmate de beoordelaar meer annotaties schrijft, is deze positiever over een te beoordelen annotatie.

Andere effecten zijn niet gevonden, maar dat betekent niet dat ze er niet zijn. Met name voor leeftijd ligt het, gezien de lage respons en het relatief sterke effect dat is gevonden, voor de hand dat zij een rol zou kunnen spelen bij de beoordeling van annotaties. Het valt niet precies te duiden wat leeftijd in deze context vertegenwoordigt. Onwaarschijnlijk is dat leeftijd te maken heeft met affiniteit met het vakgebied (hoe ouder, hoe meer affiniteit), aangezien affiniteit met het vakgebied meegenomen is, maar geen effect bleek te hebben. Bovendien correleren de twee elementen niet ($r = -.08$, *ns*). Een andere verklaring zou zijn dat naarmate men ouder wordt, men meer waardering krijgt voor annotaties. Het blijft echter speculeren, ook omdat in deze studie (nog) geen effect is gevonden van leeftijd.

De resultaten lijken in ieder geval op te gaan voor gepromoveerden die in het algemeen positief tegenover annotaties staan. Het is echter niet duidelijk of de resultaten ook kunnen worden gegeneraliseerd naar gepromoveerden die in het algemeen (heel) kritisch zijn over de kwaliteit van annotaties. Hoewel mag worden verwacht dat deze groep kritisch is over annotaties, is niet duidelijk of zij niettemin annotaties waarin implicaties worden geschetst positiever (minder negatief) beoordelen dan annotaties waarin dit element ontbreekt. Het zal moeilijk zijn om dit te toetsen. Op voorhand is het niet of moeilijk te bepalen wie annotaties negatief (minder positief) beoordelen. Daarbij komt dat het moeilijk

⁵³ De resultaten van deze analyse zijn daarom niet weergegeven. Behalve dat geen van de opgenomen concepten significante resultaten oplevert, ontstaat het probleem van multicollineariteit, met name bij de predictor ‘beschrijving’ ($VIF = 2.9$) en de predictor ‘implicaties’ ($VIF = 4.4$). Dat laatste kan overigens worden verklaard omdat de predictor ‘implicaties’ hetzelfde meet als ‘versie’.

zal zijn om deze groep bereid te krijgen om aan onderzoek naar annotaties mee te werken als zij daar het belang niet van inzien.

3 Studie 2 – Toetsen aan ‘hoger’ recht en de kwaliteit van annotaties

3.1 Onderzoeksvraag en opzet

De tweede studie richtte zich op de vraag of er een causale relatie is tussen het toetsen aan ‘hoger’ recht, dat wil zeggen beginselen of verdragsbepalingen, en de academische kwaliteit van annotaties. Op basis van de analyse zoals die is uitgevoerd in het vorige hoofdstuk, was de verwachting dat er een relatie zou bestaan tussen het kenmerk en de kwaliteit van de annotatie.

H₂: Het toetsen aan ‘hoger’ recht leidt tot een betere beoordeling van de academische kwaliteit van annotaties.

Net als in studie 1 is het aantal deelnemers verdeeld over twee groepen. De eerste groep kreeg een annotatie te lezen waarin aan ‘hoger’ recht werd getoetst. In de annotatie die aan de andere groep is voorgelegd, ontbraken de passages waarin aan ‘hoger’ recht werd getoetst. Aan de hand van verschillende stellingen over de kwaliteit van de annotatie is nagegaan of verschillen konden worden toegeschreven aan het al dan niet aanwezig zijn van het toetsen aan ‘hoger’ recht. De stellingen hadden tevens tot doel uit te sluiten dat eventuele verschillen door andere factoren zijn veroorzaakt, zoals door de structuur en de leesbaarheid.

3.2 Methode

Sample

De annotatie en de bijbehorende vragenlijst zijn in elektronische vorm verstuurd aan 110 personen. De deelnemers waren gepromoveerde juristen verbonden aan de departementen bestuursrecht⁵⁴ van de Nederlandse rechtenfaculteiten. Net als in studie 1 was het idee dat gepromoveerden binnen dit rechtsgebied in staat zouden moeten zijn om de kwaliteit van een annotatie te beoordelen. Gepromoveerden die in een eerdere fase van het onderzoek betrokken zijn, zijn niet benaderd.

⁵⁴ Bij meerdere universiteiten zijn bestuursrecht en staatsrecht ondergebracht in dezelfde eenheid (vakgroep, departement, etc.). Daarom is in de vragenlijst naar een specialisatie gevraagd.

Van de aangeschreven personen heeft 15,5% ($N = 17$) de volledige vragenlijst ingevuld. Zes respondenten hebben de vragenlijst deels ingevuld, waarvan (in volgorde van de vragenlijst) één bij het geven van de gevraagde samenvatting, één bij het beoordelen van de annotatie en vier bij of tijdens het beoordelen van de stellingen die bedoeld waren om de redenen te achterhalen waarop de respondenten hun beoordelingen hebben gebaseerd. Om statistisch significante effecten te kunnen vinden is, gezien de lage respons en de in studie 1 gevonden invloed van de predictor ‘affiniteit met schrijven annotaties’, besloten alle respondenten in de analyse te betrekken, met uitzondering van de respondenten die zijn afgehaakt bij of tijdens het beoordelen van de stellingen over hun beoordeling van de annotatie. Dit betekent dat de invloed van leeftijd, leesbaarheid/structuur/omvang, opvattingen en bekendheid met de annotator, zoals dit in studie 1 wel is gebeurd, niet is getoetst, maar wel de effecten van versie (onvolledig/volledig) en affiniteit met het schrijven van annotaties.

Een nadere inspectie leerde voorts dat één respondent geen deskundige is op het gebied van het bestuursrecht (‘Ik ben civilist, al staat mijn leerstoel dicht bij bestuursrecht’). Deze respondent is buiten de analyse gelaten. Datzelfde geldt voor de respondent die de onvolledige versie had gekregen, maar zelf de originele versie erbij heeft gepakt:

‘[...] heb ik gezocht naar de gepubliceerde noot. Deze blijkt af te wijken van het pdf bestand. Tekortkomingen die ik signaleerde in deze uitspraak zag ik daar niet (wel literatuurverwijzingen, bespreking strafrecht en [...], meer rechtspraak). Een beduidend meer wetenschappelijke noot [dan] de pdf versie die ik heb gelezen.’

Dit brengt het aantal respondenten dat in de analyses is meegenomen op 19.

Van deze respondenten was 20% ($N = 3$) vrouw en 80% ($N = 12$) man.⁵⁵ In termen van aanstelling bestond de groep respondenten uit 6 voltijds hoogleraren, 1 bijzonder hoogleraar, 1 emeritus hoogleraar, 3 universitair hoofddocenten en 4 universitair docenten.⁵⁶ Van hen zijn er negen vóór het jaar 2000 gepromoveerd, de andere zes in het jaar 2000 of daarna.⁵⁷ De gemiddelde leeftijd van de respondenten was 47 jaar.

De annotatie die is gekozen, is annotatie nr. 3, die door beide beoordelaars als goed/uitstekend is aangemerkt. Ongeveer twee vijfde van de noot bestaat uit het toetsen aan ‘hogere’ recht. Voor het overige bestaat de noot uit het verklaren van de uitspraak, het constateren van een motiveringsgebrek, het vergelijken met eerdere jurisprudentie, het schetsen van de implicaties en uit een

⁵⁵ Van vier respondenten was het geslacht onbekend.

⁵⁶ Van vier respondenten was de aanstelling onbekend.

⁵⁷ Van vier respondenten was het promotiejaar onbekend.

weergave van de feiten, de uitkomst en het geldende recht dat op de uitspraak van toepassing is. De andere annotaties waarin werd getoetst aan ‘hoger’ recht, bleken minder geschikt. Voor annotatie nr. 6 geldt dat indien de passages waarin aan ‘hoger’ recht wordt getoetst, worden weggelaten, de passages daarna los komen te staan van het overige deel van de annotatie. Voor annotatie nr. 9 geldt dat het toetsen aan ‘hoger’ recht slechts een klein deel van de noot (één korte alinea) beslaat, dat het schetsen van implicaties dominant is binnen de betreffende noot en dat daarom niet zeker is dat (alleen) het toetsen aan ‘hoger’ recht de positieve waarderingen verklaart.

Procedure

Aan de helft van de deelnemers is een annotatie voorgelegd waarin de annotator toetst aan ‘hoger’ recht. De andere helft kreeg een annotatie voorgelegd waarin passages met dit kenmerk waren weggelaten. De deelnemers, die random over de groepen zijn verdeeld, dienden de annotatie te lezen en vervolgens de vragenlijst in te vullen.

Evenals in studie 1 is aan de deelnemers, nadat zij de annotatie hadden doorgenomen, gevraagd om een kwaliteitsoordeel te geven over de annotatie die zij voorgelegd hebben gekregen. Zij dienden hun oordeel tot uitdrukking te brengen op een Likertschaal (zevenpuntschaal), waarna zij zijn verzocht een hoeveelheid stellingen te beoordelen, eveneens op een zevenpuntschaal. Ook hier hadden de stellingen betrekking op de/het lengte/leesbaarheid/taalgebruik, het bronnengebruik, de relevantie van de annotatie, het weergeven van de casus, de uitspraak en op het geldende recht dat op het geval van toepassing was. Verder zijn stellingen opgenomen over het ontwikkelen van een criterium en een stelling over de manipulatie (de weggelaten passages). De stellingen zijn met name opgenomen om te kunnen toetsen of eventuele verschillen tussen de ‘volledige’ noot en de ‘gecensureerde’ noot kunnen worden toegeschreven aan het toetsen aan ‘hoger’ recht, of dat andere verklaringen een rol spelen.

3.3 Resultaten

Van de respondenten die zijn geanalyseerd, hebben er 10 de volledige annotatie voorgelegd gekregen en hebben 9 respondenten de onvolledige versie beoordeeld. Eerstgenoemden beoordeelden de annotatie gemiddeld met een score van 4.67 ($SD = 1.00$), laatstgenoemden gemiddeld met een 5.80 ($SD = 1.14$). Hoewel de scores zijn gemeten op een zevenpuntschaal, is de schaal feitelijk een vijf-puntschaal, aangezien de laagste score een ‘3’ is en de hoogste score een ‘7’.

Het gevonden verschil is in lijn met de verwachting dat het toetsen aan ‘hoger’ recht de beoordeling van annotaties in positieve mate beïnvloedt. Om vast te stellen dat de kans klein is dat het gevonden verschil op toeval berust, is

een reeks van regressieanalyses uitgevoerd (Tabel 10). In model 1 is getoetst of de versie (onvolledig/volledig) een effect heeft op de beoordeling. Model 2 gaat na of het effect ook wordt gevonden indien rekening wordt gehouden met verschillen in affiniteit met het vakgebied/onderwerp en affiniteit met het schrijven van annotaties. In studie 1 bleek het eerste element niet, maar het laatste element wel van invloed op de beoordeling. In model 3 zijn ook de predictoren ‘opvattingen over annotaties’, bekendheid met de annotator en leeftijd toegevoegd. Bij de vier respondenten die geen scores hadden op deze predictoren, zijn de ontbrekende waarden vervangen door het gemiddelde. Hierdoor blijft het aantal respondenten dat is geanalyseerd, gelijk. Voorts is de betrouwbaarheid gemeten van de zeven items waarmee het concept ‘affiniteit met vakgebied’ is gemeten. De betrouwbaarheid van deze schaal bleek hoog (Cronbach’s Alpha = .92). Dat gold, in tegenstelling tot studie 1, niet voor de items over ‘opvattingen over annotaties’ (Cronbach’s Alpha = .30).⁵⁸ Er is dan ook besloten om één van deze items mee te nemen, namelijk de stelling die luidt: ‘Gemiddeld genomen beoordeel ik de kwaliteit van annotaties als ... (1=zwak, 7=uitstekend)’. Dit item staat in Tabel 11 opgenomen als ‘opvatting over annotaties’. De gemiddelde score voor dit item is 4.87 ($SD = .92$), gemeten op een zevenpuntschaal. De andere stelling die de opvattingen van respondenten over annotaties in het algemeen toetste (‘Annotaties dienen bij onderzoeksbeoordelingen als academische publicatie te worden aangemerkt (1=helemaal mee oneens, 7=helemaal mee eens)’), zij het minder direct dan de stelling die in onderstaande tabel is opgenomen, gaf een gemiddelde score van 4.67 ($SD = 1.72$), eveneens gemeten op een zevenpuntschaal.

⁵⁸ Het is mogelijk dat het verschil met studie 1, waar de betrouwbaarheid van dezelfde schaal wel voldoende groot was, het gevolg is van het lagere aantal respondenten ($N = 15$) in studie 2.

Tabel 11 – Resultaten regressieanalyse

Variabele	Model 1		Model 2		Model 3	
	B (s.e.)	β	B (s.e.)	β	B (s.e.)	β
Constant	3.533 (.792)	-	2.687 (1.525)	-	1.484 (2.117)	-
Versie (onvolledig/volledig)	1.133* (.493)	.49	1.163† (.582)	.50	.570 (.539)	.25
Bekendheid met annotator (nee/ja)					.340 (.634)	.10
Affiniteit vakgebied			.104 (.253)	.11	.201 (.222)	.21
Affiniteit schrijven annotaties			.204 (.270)	.18	.356 (.264)	.32
Opvatting over annotaties					.663* (.307)	.45
Leesbaarheid/structuur/omvang (perceptie)					-.503* (.217)	-
Leeftijd					-.011 (.025)	-
R²	.24		.29		.66	
F	5.28		2.08		3.05	
N	19		19		19	

Noot: * $p < .05$. † $p < .10$.

De resultaten van model 1 laten zien dat de respondenten die de volledige versie van de annotatie hebben gelezen, de annotatie positiever hebben beoordeeld dan de respondenten die de niet-volledige versie hebben gelezen ($\beta = .49$, $p = .04$). Ook model 2, waar rekening is gehouden met de invloed van ‘affiniteit vakgebied’ en ‘affiniteit met schrijven annotaties’, geeft een aanwijzing voor dit effect ($\beta = .50$, $p = .06$).

De modellen maken echter niet onmiddellijk duidelijk dat de positievere beoordeling veroorzaakt wordt door de aan- of afwezigheid van het kenmerk ‘toetsen aan hoger recht’. Het effect zou immers ook kunnen worden verklaard door aspecten van leesbaarheid, omvang en structuur (de onvolledige annotatie is te klein qua omvang, heeft een gebrekkige structuur, etc.). Net als in studie 1 is aan de hand van zeven stellingen gemeten hoe de leesbaarheid/structuur/omvang is gepercipieerd door de respondenten. Deze schaal bleek echter onvoldoende betrouwbaar (Cronbach’s Alpha = .54).⁵⁹ Na het verwijderen van vier items bleek de schaal echter wel een (voldoende) hoge betrouw-

⁵⁹ Net als bij andere schalen is het mogelijk dat het verschil met studie 1, waar de betrouwbaarheid van dezelfde schaal wel voldoende groot was, het gevolg is van het lage aantal respondenten ($N = 15$) in studie 2 met een score voor de items van deze schaal.

baarheid op te leveren (Cronbach's Alpha = .76). Het concept maakt onderdeel uit van model 3.

Model 3 is getoetst aan de hand van een regressieanalyse die vergelijkbaar is met model 2, maar waar predictoren aan zijn toegevoegd. Omdat vier respondenten geen scores hadden op de predictoren 'bekendheid met annotator', 'opvatting over annotaties', 'leesbaarheid/structuur/omvang' en 'leeftijd', zou dit betekenen dat deze respondenten niet in de analyse zouden worden meegenomen. Dit zou het aantal te analyseren personen op 15 brengen, wat gevolgen heeft voor de betrouwbaarheid ('power') van de analyse. Er is daarom voor gekozen om de ontbrekende waarden te vervangen door het groepsgemiddelde. Overigens lost dit het 'power'-probleem niet geheel op: het aantal geanalyseerde respondenten ($N = 19$) blijft laag, wat de kans verkleint op het verkrijgen van statistisch significante resultaten.

Model 3 laat effecten zien voor 'leesbaarheid/structuur/omvang' ($\beta = -.51, p = .04$) en voor 'opvatting over annotaties' ($\beta = .45, p = .05$), maar niet voor versie (onvolledig/volledig) ($\beta = .25, p = .31$).

3.4 Discussie

De omstandigheid dat model 3 effecten geeft voor 'leesbaarheid/structuur/omvang' en voor 'opvatting over annotaties', maar niet voor versie (H_2), suggereert dat het verschil in beoordeling in ieder geval kan worden verklaard door aspecten van leesbaarheid/structuur/omvang (en de opvattingen van respondenten over annotaties). Zij betekent echter niet zonder meer dat er geen effect van het toetsen aan 'hoger' recht is. Afgaand op de relatief hoge *effect size* ($\beta = .25$) en het lage aantal respondenten, valt te verwachten dat er ook voor het toetsen aan 'hoger' recht significante effecten zouden worden gevonden indien het aantal respondenten groter zou zijn en zij in dezelfde lijn zouden hebben geantwoord als de onderzochte respondenten. Hetzelfde geldt overigens voor de predictoren 'affiniteit met het vakgebied' ($\beta = .21$) en 'affiniteit schrijven annotaties' ($\beta = .32$).

Interessant is verder dat de predictor 'opvattingen' in deze studie een effect blijkt te hebben op de beoordeling van annotaties. Hoewel dit intuïtief niet verbaast – een negatief beeld van annotaties in het algemeen beïnvloedt het oordeel van een individuele annotatie – werd in studie 1 geen effect gevonden. Een verklaring, anders dan dat het gevonden effect te maken heeft met de respondenten die hebben deelgenomen aan deze studie, is niet onmiddellijk voorhanden. Wel blijkt het belangrijk te zijn dat dit element wordt meegenomen bij analyses over het beoordelen van annotaties.

De slotsom van deze studie is dat er aanwijzingen, maar geen harde bewijzen zijn gevonden voor het effect van het 'toetsen' aan hoger recht. Dat geen

harde bewijzen zijn gevonden, lijkt in belangrijke mate te maken te hebben met het lage aantal respondenten dat heeft meegedaan. Het vergroten van de respons is echter lastig, gezien de omstandigheid dat de tijdsinvestering relatief groot is (ca. 45-60 min. per respondent) en de respondenten drukbezette personen zijn die bovendien niet onmiddellijk een eigen belang hebben om aan het onderzoek mee te doen.

4 Conclusie

De voorspelde positieve invloed van het schetsen van implicaties en het toetsen aan ‘hoger’ recht op de beoordeling van annotaties wordt bevestigd door de resultaten uit de studies. Studie 1 laat zien dat het schetsen van implicaties van de uitspraak in een annotatie leidt tot een betere beoordeling van een annotatie (H_1), onder constanthouding van affiniteit met het vakgebied, affiniteit met het schrijven van annotaties, opvatting over annotaties, bekendheid met de annotator, leeftijd en leesbaarheid/structuur/omvang. In studie 2 werd vastgesteld dat er aanwijzingen, maar geen harde bewijzen zijn gevonden voor de positieve invloed van het ‘toetsen’ aan hoger recht op de beoordeling van annotaties (H_2). Bij een hoger aantal respondenten zou, gezien de gevonden *effect sizes*, het bewijs waarschijnlijk wel ‘hard’ zijn.

Meer in het algemeen geldt dat de resultaten uit de studies de resultaten bevestigen van de analyse zoals die eerder is uitgevoerd en waarin annotaties die als goed/uitstekend zijn aangemerkt zijn vergeleken met annotaties die als matig/slecht zijn gekwalificeerd (Hoofdstuk 3). Deze combinatie maakt dat er sterke aanwijzingen zijn voor het gegeven dat het schetsen van implicaties en het toetsen aan ‘hoger’ recht een positieve invloed hebben op hoe positief dan wel negatief annotaties worden beoordeeld.

Wetenschappelijkheid van annotaties

1 Inleiding

In het debat over de aard, het nut en de status van annotaties en juridisch onderzoek in het algemeen is nogal eens een beroep gedaan op de (on)wetenschappelijkheid ervan. Dit hoofdstuk staat stil bij de vraag welke betekenis de resultaten uit de vorige hoofdstukken hebben voor de mate waarin annotaties als wetenschappelijk kunnen worden aangemerkt.

Wetenschappelijkheid staat niet gelijk aan nut of relevantie, maar houdt er wel mee verband. Het predicaat ‘wetenschappelijk’ heeft belangrijke voordelen. Met name de associatie met betrouwbaarheid is van belang. Wetenschappelijke conclusies worden geacht aan een grondig selectieproces te zijn onderworpen, deze toets der kritiek te hebben doorstaan en daarmee betrouwbaar te zijn.⁶⁰ Weliswaar kunnen wetenschappelijke conclusies voor discussie vatbaar zijn, later worden bijgesteld, worden genuanceerd of zelfs worden verworpen, maar dit neemt niet weg dat er op mag worden vertrouwd dat de resultaten van wetenschappelijk onderzoek voldoen aan een bepaalde minimumstandaard van kwaliteit en betrouwbaarheid. Tegelijk staat wetenschappelijkheid los van de begrippen nut en relevantie. Net zoals een arts, een loodgieter of een journalist zinvol werk kan leveren, kan ook een annotator stukken schrijven die relevant zijn voor bijvoorbeeld rechters of advocaten zonder dat deze stukken wetenschappelijk zijn.

Het zal blijken dat het onderscheid tussen wetenschappelijk en niet-wetenschappelijk minder voor de hand ligt dan wel eens wordt gepresenteerd in publicaties over de wetenschappelijkheid van juridisch onderzoek (par. 2).⁶¹ Om teleurstellingen te voorkomen: het zal niet mogelijk zijn om concrete criteria aan te reiken aan de hand waarvan publicaties, annotaties in het bijzonder, kunnen worden aangemerkt als wetenschappelijk of niet-wetenschappelijk, of aan de hand waarvan overeenkomsten kunnen worden gevonden tussen de gevonden criteria en de eisen die aan wetenschappelijke publicaties worden gesteld. De

⁶⁰ Hansson, S.O. (2003). ‘Science and Pseudo-Science.’ The Stanford Encyclopedia of Philosophy Winter 2003 Edition. Retrieved 17 March 2011, from <http://plato.stanford.edu/entries/pseudo-science/>. Hoe dat beeld is ontstaan, is niet eenduidig te verklaren. Verschillende sociale en historische omstandigheden spelen vermoedelijk een rol, waaronder de gepercipieerde vooruitgang die bepaalde typen onderzoek (met name de bètawetenschappen) hebben geboekt.

⁶¹ Ik beperk mij tot Nederlandstalige literatuur.

publicaties, opvattingen en theorieën waarnaar ik hierna verwijs, zijn veelal niet geschreven om concrete criteria aan te reiken, of een beoordelingsinstrument te ontwikkelen om een wetenschappelijke (juridische) publicatie te beoordelen op de mate van wetenschappelijkheid, laat staan dat zij zijn toegespitst op annotaties.⁶²

Dat er geen concrete criteria bestaan om publicaties te beoordelen, is een belangrijke vaststelling. Dit wil echter niet zeggen dat wetenschappelijk onderzoek geen kenmerken heeft die zicht kunnen geven op het wetenschappelijke karakter van annotaties. Het zal blijken dat de wetenschappelijkheid van annotaties genuanceerder ligt dan soms wel wordt voorgesteld.

Dit hoofdstuk tracht deze kenmerken te achterhalen en in verband te brengen met annotaties en de kwaliteit ervan. Daartoe wordt, kort, het debat geschetst over de wetenschappelijkheid van onderzoek zoals die zich binnen de wetenschapsfilosofie heeft afgespeeld, te beginnen met het kritisch-rationalisme van Popper (par. 3), gevolgd door Kuhns paradigmatheorie (par. 4), Lakatos' theorie van onderzoeksprogramma's (par. 5) en de multicriteria-benadering die sommigen hebben beschreven (par. 6). In een tussenstand wordt bezien welke kenmerken van wetenschap het debat opleveren (par. 7) en of deze kenmerken terug te vinden zijn in de (verschillende typen) annotaties die eerder in deze studie zijn onderzocht (par. 8-9).

2 Wetenschappelijkheid van juridisch onderzoek

De vraag naar de wetenschappelijkheid van juridisch onderzoek is niet nieuw. Zo schreef Kranenburg in 1955, naar aanleiding van een 20 jaar daarvoor verschenen publicatie:⁶³

‘Over aard en wezen der rechtswetenschap en over haar methoden bestaat, ondanks haar ouderdom, groot meningsverschil. In de laatste decennia meer dan ooit; is zij wel een “wetenschap”? zo wordt gevraagd.’⁶⁴

Hij beschouwde het ordenen, systematiseren en kritisch analyseren van verschijnselen als functie van de rechtswetenschap, met als doel het begrijpen van het recht als samenhangend en verstaanbaar geheel.⁶⁵

‘Rechtsvorming gaat aan rechtswetenschap vooraf, maar de drang van den menselijken geest naar begrip, verklaring, doet, zodra de rechtsvorming een zekere ontwikke-

⁶² Sommige wetenschapsfilosofen benadrukken dit ook, zie bijvoorbeeld Kuhn 1977, p. xxi-xxii.

⁶³ Lundstedt 1930/1936.

⁶⁴ Kranenburg 1955, p. 1.

⁶⁵ Kranenburg 1955, p. 23.

ling heeft bereikt, de wetenschap ontstaan. Ordenend, systematiserend, verklarend werkt het denken ook hier, evenals op het gebied van andere verschijnselencomplexen.⁶⁶

In meer recente literatuur beroepen auteurs zich op wetenschapsfilosofen van na 1955 wanneer zij de rechtswetenschappelijkheid van juridisch onderzoek aanvallen of bepleiten. Een uitgewerkte, systematische toepassing van die opvattingen op juridisch onderzoek ontbreekt echter meestal. Hierna vat ik kort samen de werken van auteurs die wel uitgebreider en op een meer systematische wijze de wetenschappelijkheid van juridisch onderzoek of bepaalde vormen ervan hebben bepleit. Ik richt mij op Nederlandstalige publicaties over de afgelopen vijf jaar, omdat die gaan over recent en/of actueel juridisch onderzoek.

Wendt schreef over het toetsen van hypothesen in het kader van tekstinterpretatie.⁶⁷ Hij beschouwt de, wat hij aanduidt als, kritische hypothetisch-deductieve methode als een wetenschappelijke methode van tekstinterpretatie. Onder verwijzing naar Popper omschrijft Wendt de kritische hypothetisch-deductieve methode in termen van een onderzoeker die tijdens het lezen hypothesen opstelt over de betekenis van de tekst en deze vervolgens toetst door zich aan de hand van de tekst af te vragen welke interpretatie het meest verdedigbaar is.⁶⁸ De onderzoeker gaat daarbij op zoek naar vindplaatsen in een tekst die zijn hypothese(n) ondersteunen of aanleiding geven om deze te verwerpen. Het is essentieel om daarbij kritisch te werk te gaan, aldus Wendt, want het risico bestaat dat een hypothese op te lichtvaardige gronden wordt bevestigd of verworpen. Intuïtie en creativiteit zouden daarbij een belangrijke rol vervullen.

Wendt gebruikt ter illustratie de uitleg van overeenkomsten. Hij beschrijft een interpretatiekader, alsook de totstandkoming ervan, dat kan worden gebruikt om te beoordelen wanneer een overeenkomst tekstueel moet worden uitgelegd. Het vertrekpunt hierbij is de gedachte of aanname dat partijen zichzelf niet tekort willen hebben doen en als dat wel het geval is, dat zij daar geen misbruik van mogen maken.⁶⁹ Dit voert naar een tekstuele uitleg, uitgaande van een overeenkomst tussen twee partijen. Ook die tekstuele uitleg is vaak gebaseerd op een interpretatiekader, bestaande uit verschillende aannames die op zichzelf voor verwerping vatbaar zijn. Een voorbeeld is de veronderstelling dat het de bedoeling van auteurs is om een consistente tekst samen te stellen, dat wil zeggen, een tekst die één geheel is. Hieruit zijn verschillende uitlegregels te distilleren, zoals: alle gegevens uit de tekst zijn in beginsel relevant voor de interpretatie, verwante tekstgedeelten zijn van belang voor de uitleg van een bepaalde passa-

⁶⁶ Kranenburg 1955, p. 22.

⁶⁷ Wendt 2008, p. 75 e.v.

⁶⁸ Wendt 2008, p. 77.

⁶⁹ Wendt 2008, p. 80.

ge, ogenschijnlijke tegenstrijdigheden zijn onwaarschijnlijk en minder duidelijke passages dienen te worden geïnterpreteerd in het licht van duidelijkere passages.⁷⁰

Een interpretatiekader zoals Wendt dat omschrijft, dient als een zeef om de relevante feiten te selecteren en om te toetsen of de feiten aansluiten bij het interpretatiekader en wat dit beoogt. De onderzoeker bepaalt, met andere woorden, aan de hand van welke beginselen, regels of gezichtspunten een overeenkomst moet worden uitgelegd, waarna wordt bekeken of de feiten hieraan voldoen. De confrontatie met concrete gevallen maakt het mogelijk om het interpretatiekader te toetsen: leidt de toepassing ervan tot redelijke uitkomsten? Is dat niet het geval, dan wordt het interpretatiekader aangepast en wordt de houdbaarheid opnieuw bekeken. Dit proces blijft doorgaan tot het moment dat een redelijke uitkomst wordt gevonden. Of een uitkomst redelijk is, hangt volgens Wendt af van een hoeveelheid van belangen, beginselen en gezichtspunten, waaronder de maatschappelijke aanvaardbaarheid, de effecten op het gedrag van personen en op de kosten die als gevolg van de beslissing moeten worden gemaakt, de consistentie van de beslissing met het oog op gevallen die zich in de toekomst voor zouden kunnen doen, alsook de verhouding tot eerdere jurisprudentie (is de beslissing daarmee in lijn?) en het stelsel van de wet. Voorts is van belang de mate waarin welk element wordt meegewogen.

Hesselink benadert de wetenschappelijkheid op een andere manier. Hij stelt dat juridisch onderzoek geen hypothesen produceert die empirisch kunnen worden getoetst en dat het daarom niet wetenschappelijk kan zijn volgens Poppers falsificatiecriterium.⁷¹ Juristen verdedigen in de regel de eigen theorie of opvatting in plaats van deze aan te vallen.⁷² Onder verwijzing naar, met name, Bruno Latour⁷³ stelt Hesselink dat de kwalificatie ‘wetenschappelijk’ geconstrueerd wordt. Wat anderen doen met onderzoeksresultaten bepaalt of resultaten voor ‘waar’ of feit worden aangenomen.⁷⁴ De verschillen tussen wetenschapsdisciplines in manieren van het doen van onderzoek zijn daarom veel meer graduëel en argumentatief dan categorisch.⁷⁵ In essentie, aldus Hesselink, verschillen de wijzen waarop kennis wordt vergaard niet tussen de juridische discipline en ‘hard sciences’.

⁷⁰ Wendt 2008, p. 78-79, alsook noot 72 aldaar.

⁷¹ Hesselink 2009, p. 24.

⁷² Hesselink 2009, p. 24-25.

⁷³ Latour 1987.

⁷⁴ Hesselink 2009, p. 27.

⁷⁵ Hesselink 2009, p. 27. Deze conclusie lijkt overigens niet direct te kunnen worden afgeleid uit Latour 1987, waarin kort gezegd vooral een visie wordt gepresenteerd over hoe (wetenschappelijke) kennis tot stand komt en wanneer zij wordt geaccepteerd, en waarin het onderscheid wetenschappelijk/niet-wetenschappelijk minder aan de orde is.

Smits, tot slot, bepleit een, wat hij noemt, empirisch-normatieve methode als voorbeeld voor rechtswetenschappelijk onderzoek.⁷⁶ Deze methode komt kort gezegd neer op het vergelijken van argumenten die in rechtsstelsels worden gebruikt om bepaalde gevallen te beslechten waarin sprake is van conflicterende belangen. Het idee hierachter is dat rechtsstelsels als laboratoria kunnen worden gezien waarin oplossingen zijn bedacht voor het omgaan met bepaalde typen conflicten. Volgens Smits voldoet deze methode aan de eisen die aan een wetenschap worden gesteld: (1) het systematiseren van kennis, dat wil zeggen een beschrijving, evaluatie of verklaring vanuit een bestaand kader, (2) het verwerven van kennis via een door de academische gemeenschap erkende onderzoeksmethode of via een als zodanig erkende praktijk en (3) het streven naar universele kennis, dat wil zeggen kennis die beoogt resultaten te veralgemeniseren.⁷⁷ Smits maakt niet duidelijk waar hij deze eisen vandaan haalt.

3 Kritisch rationalisme

Het kritisch rationalisme, zoals verdedigd door Popper, bouwde voort op het positivisme, dat tegenwicht bood aan een samenleving waarin kerk en staat als autoriteiten golden. Het probleem met het positivisme was volgens Popper vooral het verificatiecriterium. Het viel hem op dat bepaalde theorieën, zoals die van Marx en Freud, vrijwel alles konden verklaren.⁷⁸ Zij leken altijd zo te kunnen worden aangepast dat ogenschijnlijke afwijkingen er toch een plaats in konden krijgen. Popper plaatste hier het voorbeeld van de relativiteitstheorie van Einstein tegenover.⁷⁹ Die zou eenvoudig kunnen worden weerlegd als zou blijken dat bepaalde sterren niet van positie veranderen of als de verplaatsing van die ster niet overeen zou komen met wat Einstein had voorspeld. De theorie van Einstein had daarmee iets wat de theorieën van Marx en Freud niet hadden. Daar waar laatstgenoemde theorieën altijd juist waren, droeg Einsteins theorie het risico in zich om verworpen te worden.

In dit onderscheid zocht Popper het kenmerk voor wetenschappelijkheid, waarbij wetenschappelijkheid overigens niet moet worden verward met ‘juist’ of ‘waarheid’. Het zoeken naar bewijs dat de eigen gedachte of theorie ondersteunt, volstond volgens hem niet.⁸⁰ Voor Popper was de mogelijkheid om een theorie te kunnen verworpen essentieel voor de wetenschappelijkheid ervan. De wetenschappelijkheid van onderzoek was dan ook niet gelegen in verificatie, maar in

⁷⁶ Smits 2009, p. 115.

⁷⁷ Smits 2009, p. 66-70.

⁷⁸ Popper 1974, p. 35 en p. 37-39; Popper 1983, p. 163-174; McGrew & Alspector-Kelly & Allhoff 2009, p. 459. Zie ook Popper 1980, p. 82.

⁷⁹ Popper 1974, p. 35-36; McGrew & Alspector-Kelly & Allhoff 2009, p. 459-460.

⁸⁰ Popper 1980, p. 27-48.

falsificatie.⁸¹ Indien een theorie pogingen tot falsificatie succesvol doorstaat, is zij voorlopig aanvaard, want later kan blijken dat er toch bewijs bestaat tegen de theorie. Voorts hangt de wetenschappelijkheid van een theorie af van de mate waarin zij specifiek, informatief en generaliseerbaar is.⁸²

Er moet echter worden bedacht dat het falsificatiecriterium, ook volgens Popper zelf, zich niet zonder meer leent om onderzoek en theorieën als wetenschappelijk of niet-wetenschappelijk te kwalificeren.⁸³ Dat was ook niet het doel van Popper. Zijn doel was om te komen tot een criterium aan de hand waarvan in abstracto kan worden bepaald wat het onderscheid is tussen wetenschappelijk en niet-wetenschappelijk. Om uitspraken over de wetenschappelijkheid van een theorie of een hoeveelheid uitspraken te kunnen doen, is het volgens Popper nodig om de methode te beoordelen op basis waarvan de theorie tot stand is gekomen.⁸⁴

Verschillende auteurs verwijzen, de een wat uitgebreider dan de ander, naar Popper wanneer zij juridisch onderzoek verdedigen, bekritisieren of beschrijven.⁸⁵ Ik beperk mij tot een bespreking van de dissertatie van Wendt, die veruit de meest omvangrijke studie in Nederland is waarin Poppers kritisch-rationalisme wordt toegepast op juridisch onderzoek.⁸⁶ Wendt stelt dat falsificatie deels bruikbaar is in de rechtswetenschap.⁸⁷ Volgens hem kunnen bepaalde aspecten van het recht empirisch⁸⁸ worden onderzocht, behalve waar het gaat om normatieve vragen.⁸⁹ Wendt illustreert zijn betoog onder meer aan de hand van het voorbeeld van de doodstraf.⁹⁰ Daarvan zou de preventieve werking kunnen worden onderzocht (empirisch), of worden bekeken hoeveel procent van de bevolking vóór dan wel tegen de doodstraf is (empirisch), maar de vraag of de doodstraf in absolute zin wenselijk is, kan in zijn algemeenheid niet worden gefalsificeerd.

Een beroep op Popper spreekt echter niet voor zich. Zo heeft het kritisch-rationalisme onder vuur gelegen binnen de wetenschapsfilosofie, wat aanleiding heeft gegeven voor andere benaderingen en opvattingen over de wetenschappe-

⁸¹ Popper 1980, p. 40; Popper 1974, p. 36-37; Popper 1983.

⁸² Over gradaties van falsificeerbaarheid, zie Popper 1980, p. 112-135.

⁸³ Popper 1980, p. 81-82.

⁸⁴ Popper 1980, p. 82.

⁸⁵ Zie onder meer Wendt 2009; Smith e.a. 2008; Wendt 2008; Degenkamp 2004; Tjittes 2005; Oosten 2004; Franken 2004; Stolker 2004.

⁸⁶ Wendt 2008.

⁸⁷ Wendt 2008, p. 64-65.

⁸⁸ Empirisch dient hier te worden opgevat in ruime zin (waarnemen van de werkelijkheid) en niet in termen van 'toepassen van sociaalwetenschappelijke of natuurwetenschappelijke' onderzoeksmethoden.

⁸⁹ Wendt 2008, p. 65.

⁹⁰ Wendt 2008, p. 65.

lijkheid van onderzoek en theorieën. De kritiek valt uiteen in verschillende onderdelen.⁹¹ Een eerste bezwaar is dat als feiten niet corresponderen met de theorie, dit niet noodzakelijkerwijs aan de theorie hoeft te liggen. De manier waarop de feiten zijn verkregen kan bijvoorbeeld niet deugen. Of daar sprake van is, zal vaak moeilijk en meestal na verloop van tijd vast te stellen zijn.

Een tweede kritiekpunt op het kritisch-rationalisme is dat iedere theorie onvolkomenheden of elementen bevat die niet stroken met de werkelijkheid. Bij een strikte toepassing van de falsificatiegedachte zou dit betekenen dat vrijwel geen enkele theorie zou overleven. De werkelijkheid is echter anders. In de praktijk van onderzoek proberen onderzoekers dergelijke problemen op te lossen. De geschiedenis leert dat dit soms ook met succes gebeurt, wat dan leidt tot het verfijnen van bestaande (gefalsificeerde) theorieën.

Verder stelde Popper dat wetenschap tot doel heeft op zoek te gaan naar de waarheid, maar dat theorieën nooit ‘waar’ kunnen zijn. Kritiek hierop is dat het niet rationeel is om te streven naar iets dat nooit zal kunnen worden bereikt. Tot slot is de theorie van Popper zelf niet falsificeerbaar, nu zij een normatieve wetenschapsopvatting betreft.

Behalve de algemene kritiek op het kritisch-rationalisme, is het de vraag of het zinvol is passages uit Popper te gebruiken om de wetenschappelijkheid van juridisch onderzoek te verdedigen of beargumenteren. Zo suggereert Wendt dat Popper het wetenschappelijke proces heeft vergeleken met rechterlijke besluitvorming. Hij leidt dit af uit de vergelijking zoals Popper die maakt tussen wetenschappelijk onderzoek en beslissingen van juryleden om een verdachte al dan niet te veroordelen. In beide vormen van besluitvorming zou een ‘sprong’⁹² zitten, namelijk die tussen *truth* en *verdict*.⁹³

Het is echter de vraag of Popper hiermee heeft willen beweren dat wat rechters doen, als wetenschappelijk is aan te merken. In de passages waar Wendt naar verwijst, illustreert Popper dat de sprong van waarheid naar conclusie verloopt via een theorie, dat wil zeggen de bril die de onderzoeker respectievelijk de rechter op zet. In onderzoek stuurt een wetenschappelijke theorie de onderzoeksvragen, de hypothesen, de waarneming en de interpretatie. In een juridisch proces vormen bijvoorbeeld procedurele regels de theorie waaraan een juryuitspraak (*verdict*) wordt getoetst. Deze theorie stuurt niet alleen de vragen die moeten worden gesteld (‘is er gerede twijfel?’, ‘is de beslissing gemotiveerd?’), maar ook de criteria waar een beslissing van bijvoorbeeld een rechter of een jury aan moet worden getoetst.⁹⁴ Aan welke theorie wordt getoetst, berust op een

⁹¹ Ik baseer mij op de heldere weergave McGrew & Alspector-Kelly & Allhoff 2009, p. 461-462.

⁹² Doelend op de beroemde ‘sprong’ van Scholten.

⁹³ Zie Wendt 2008, p. 66-67.

⁹⁴ Popper 1980, p. 110.

keuze van respectievelijk de wetgever, rechter of onderzoeker.⁹⁵ Het falsificeren ziet echter op iets anders, namelijk het toetsen of de feiten in overeenstemming zijn met de theorie.

Als het maken van het onderscheid tussen de keuze voor de theorie (*justification*) en het toetsen ervan (*decision*) van belang is voor de wetenschappelijkheid van onderzoek, betekent dit dat volgens Popper niet de ‘sprong’ als kenmerk van wetenschappelijkheid moet worden gezien, maar de falsificeerbaarheid van de ‘sprong’. Dit wil zeggen dat de mogelijkheid moet bestaan dat de feiten niet in overeenstemming kunnen worden gebracht met de juridische theorie, waarvoor ten minste vereist is dat duidelijk is aan welke theoretische uitgangspunten moet worden getoetst. En of juridische theorieën altijd falsificeerbaar zijn, valt te betwijfelen. Een rechter zal bijvoorbeeld kunnen toetsen (falsificeren) of een jury zich aan de procedurele regels heeft gehouden en daar een conclusie aan kunnen verbinden (bijvoorbeeld: jury spreekt alleen een veroordeling uit bij een unanieme beslissing), maar dat kan anders liggen als hij zich in het kader van een belangenafweging beroept op bijvoorbeeld de redelijkheid en billijkheid.⁹⁶ Als iedere beslissing kan worden gerechtvaardigd met een beroep op bijvoorbeeld de redelijkheid en billijkheid, is het theoretische uitgangspunt (redelijkheid en billijkheid) niet falsificeerbaar en de methode dus niet wetenschappelijk. Dit betekent niet dat het concept van redelijkheid en billijkheid onbruikbaar is in de praktijk. Sterker nog, het kan zelfs te prefereren zijn boven regels met meer rechtszekerheid. Als toetsingskader in wetenschappelijk (juridisch) onderzoek is het echter niet of minder geschikt.

4 Kuhns paradigmatheorie

De normatieve wetenschapsopvatting zoals Popper die verdedigde is binnen de wetenschapsfilosofie minder aan de orde en is door velen zelfs verlaten. Het kritisch-rationalisme en de kritiek erop gaf aanleiding voor anderen om nieuwe, alternatieve opvattingen te ontwikkelen. De belangrijkste opvolger was de paradigmatheorie van Kuhn. Kuhn had, anders dan Popper, een meer empirische wetenschapsopvatting.⁹⁷ Kuhn verbaasde zich over het verschil in opvatting tussen natuurwetenschappers en sociale wetenschappers over wat een probleem en methode wetenschappelijk maakt, en over het gegeven dat de wetenschappelijkheid van natuurwetenschappelijk onderzoek veel minder ter discussie stond dan de wetenschappelijkheid van sociaalwetenschappelijk onderzoek.⁹⁸

⁹⁵ Popper 1980, p. 109.

⁹⁶ Dit is precies het onderscheid dat ook Popper maakt, Popper 1980, p. 110.

⁹⁷ Hoewel deze, zoals Kuhn zelf aangeeft, soms ook normatieve opvattingen bevat, zie Kuhn 1962, p. 8.

⁹⁸ Kuhn 1962, p. x.

In zijn beroemde werk *The Structure of Scientific Revolutions* vermoedde Kuhn dat binnen disciplines waar niet of nauwelijks discussie bestaat over hoe onderzoek moet worden uitgevoerd en over de wetenschappelijkheid van het onderzoek, onderzoekers op het verkeerde been zijn gezet door de historie.⁹⁹ De relatief grote mate van overeenstemming vloeide volgens Kuhn voort uit het gegeven dat onderzoekers zijn opgeleid binnen een bepaalde onderzoekscultuur, waarbinnen het normaal is om onderzoek op een bepaalde manier te verrichten. Docenten en studieboeken hebben dit de studenten, en dus de huidige onderzoekers, geleerd. Het studiemateriaal vertelt echter meestal niet dat de bestaande manier van het doen van onderzoek vroeger anders was. Voor zover het gebeurt, wordt het gepresenteerd als een verkeerd idee dat is ingehaald door de superieure methode die vandaag de dag geldt.¹⁰⁰ Dit vormde aanleiding voor Kuhn om op zoek te gaan naar onderzoek uit het verleden, waarbij hij trachtte te verklaren hoe wetenschappelijke theorieën zijn ontstaan en hoe zij hebben kunnen overleven.

Kuhn onderscheidt in dat verband verschillende fases van wetenschap, te beginnen met die van zogenaamde *pre-paradigm period*. Hierin begint een hoeveelheid personen een verschijnsel of gebeurtenis te onderzoeken, waarbij de onderzoekers starten met verschillende opvattingen en accenten over wat moet worden onderzocht en hoe dit dient te gebeuren.¹⁰¹ Op een zeker moment zal een groep(je) onderzoekers, aldus Kuhn, erin slagen om een theorie te ontwikkelen die consistent is en die bovendien zinvolle resultaten en verklaringen produceert, dat wil zeggen, resultaten die door anderen als relevant worden beschouwd.¹⁰² De ontwikkelde theorie blijkt dan, kort gezegd, probleemoplossend te zijn en zal daarom een grotere groep onderzoekers aantrekken. Wil een theorie overleven, dan dient zij het beter te doen dan concurrerende theorieën; andere 'scholen' zullen dan uiteindelijk verdwijnen.¹⁰³

Als een dergelijke situatie ontstaat, is sprake van een, wat Kuhn noemt, normale wetenschap.¹⁰⁴ Er geldt een paradigma dat bepaalt wat wordt waargenomen, hoe dit gebeurt en hoe de resultaten behoren te worden geïnterpreteerd.¹⁰⁵ Studenten worden binnen de gangbare theorieën opgeleid, passen deze toe op bestaande situaties en als zij dit voldoende frequent hebben gedaan, op situaties waar de uitkomst nog niet van bekend is.¹⁰⁶ Onderzoekers hoeven niet

⁹⁹ Kuhn 1962, p. 1.

¹⁰⁰ Kuhn 1962, p. 1, 143.

¹⁰¹ Kuhn 1962, p. 4, 17, 47-48.

¹⁰² Kuhn 1962, p. 10-11, 17.

¹⁰³ Kuhn 1962, p. 17-19, 23.

¹⁰⁴ Kuhn 1962, p. 10.

¹⁰⁵ Kuhn 1962, p. 27.

¹⁰⁶ Zie ook Kuhn 1962, p. 5.

alles wat zij in hun onderzoek doen uit te leggen en te verantwoorden, want methoden en werkwijzen zijn algemeen aanvaard binnen de discipline.¹⁰⁷ Sterker nog, zolang onderzoekers het eens zijn over hoe onderzoek te doen, kan een discipline zonder geëxpliciteerde methodologische spelregels.¹⁰⁸

Kuhn meende dat de werkelijkheid slechts vanuit een bepaald perspectief kan worden onderzocht. Het paradigma bepaalt immers wat onderzocht moet worden en hoe dat moet gebeuren. Dit paradigma kan zelfs zo sterk dat het doel van wetenschappers niet zo zeer is om nieuwe theorieën te ontdekken, maar dat het de praktijk heeft van het toepassen van de theorie op een hoeveelheid situaties, hetgeen Kuhn aanduidt als ‘puzzelen’.¹⁰⁹ Deze praktijk leidt ertoe dat wetenschappers geneigd zijn vragen die niet binnen het bestaande paradigma kunnen worden beantwoord, alsook andere manieren van onderzoek, te bekritisieren of vrijwel onmiddellijk van de hand te wijzen.¹¹⁰

Normale wetenschap kan echter veranderen in een crisis of revolutie. Dit gebeurt als er feiten worden gevonden die afwijken van wat de theorie voorspelt. Het zal daarbij gaan om afwijkingen die de kern van de theorie raken, want voor kleine afwijkingen kunnen ad-hoc-oplossingen worden gevonden.¹¹¹ De weerstand om de theorie omver te werpen zal echter groot zijn.¹¹² Als deze afwijkingen (anomalieën) aanhouden, niet slechts ‘puzzels’ zijn en dus niet door de theorie kunnen worden verklaard, zal de vraag rijzen of de theorie zelf wel deugt.¹¹³ Wetenschappers zullen dan, zoals ook deze studie daar een voorbeeld van is, binnen de filosofie op zoek gaan naar antwoorden.¹¹⁴ Ook is de kans groot dat zij van mening zullen verschillen over welke aannames binnen de theorie onjuist zijn. Versplintering ligt dan op de loer. In dat geval gaat de discipline een onzekere periode tegemoet, eentje van crisis, een situatie die vergelijkbaar is met *pre-science*.¹¹⁵ Symptomen voor een crisis zijn volgens Kuhn: een uiting van onvrede over de bestaande manier van onderzoek, de wil om een nieuwe aanpak of benadering te proberen, het bestaan van concurrerende gedachten over hoe verder te gaan, het zoeken naar antwoorden binnen de filosofie en debatten over de kern van het onderzoek dat binnen de theorie of discipline wordt gedaan.¹¹⁶

¹⁰⁷ Kuhn 1962, p. 19-20.

¹⁰⁸ Kuhn 1962, p. 47.

¹⁰⁹ Kuhn 1962, p. 24, 35-38, 52.

¹¹⁰ Kuhn 1962, p. 24. Vergelijk uitspraken als ‘wij zijn juristen, geen sociale wetenschappers’.

¹¹¹ Kuhn 1962, p. 78, 82.

¹¹² Kuhn 1962, p. 65.

¹¹³ Kuhn 1962, p. 82. Zie ook p. 6.

¹¹⁴ Kuhn 1962, p. 87.

¹¹⁵ Kuhn 1962, p. 67-68, p. 84-85.

¹¹⁶ Kuhn 1962, p. 90. Het debat binnen het juridische forum over de wetenschappelijkheid ervan lijkt aan al deze kenmerken te voldoen.

Overigens wordt meestal pas veel later vastgesteld dat sprake is geweest van een crisis. De onderzoekers zijn zich daar op dat moment niet van bewust.

Een crisis gaat over in een revolutie op het moment dat een wetenschapper of een groep van wetenschappers in staat is gebleken een oplossing te formuleren voor één of meer van de gevonden afwijkingen.¹¹⁷ Een nieuw paradigma is dan geboren, terwijl het oude zal worden afgeschreven.¹¹⁸ Sommige wetenschappers zullen naar het nieuwe paradigma overstappen; anderen zullen het oude blijven verdedigen en binnen dit paradigma blijven zoeken naar oplossingen.¹¹⁹ Een dergelijke overgang (*paradigm shift*) beschouwt Kuhn als een revolutie. Welke theorie uiteindelijk overleeft, wordt volgens hem bepaald door het oplossend vermogen van de concurrerende paradigma's. Er bestaat geen absoluut criterium of bewijs om het oplossende vermogen of de 'betere' theorie te bepalen, maar dit neemt niet weg dat wetenschappers argumenten kunnen formuleren, overtuigd raken en daarom van mening kunnen veranderen.¹²⁰ Dat zal in het bijzonder het geval zijn als het nieuwe paradigma één of meer belangrijke problemen oplost wat onder het oude paradigma niet mogelijk was, en als het nieuwe paradigma ook een relatief groot deel van de problemen kan blijven oplossen die ook binnen het oude paradigma konden worden opgelost.¹²¹

De verschillen tussen Kuhn en Popper zijn groot. Volgens Kuhn gebeurt het verwerpen van een theorie niet via falsificatie, maar slechts onder de voorwaarde dat er een concurrerende theorie bestaat die een groter probleemoplossend vermogen heeft.¹²² Een theorie kan, met andere woorden, niet worden verworpen zonder dat een andere theorie wordt aanvaard. In het geval van een afwijking van de theorie zullen nuanceringen en ad-hoc-oplossingen binnen de bestaande theorie worden bedacht. Daarbij komt dat er geen theorie bestaat die alle afwijkingen, groot en klein, kan verklaren of oplossen.¹²³ Als alleen een belangrijke afwijking kan leiden tot het verwerpen van een theorie, is bovendien een normatief criterium nodig op basis waarvan de mate van falsificatie kan worden vastgesteld.¹²⁴ Voor een dergelijk normatief criterium geldt dat filosofen niet zoeken naar absolute criteria om de wetenschappelijkheid van een theorie vast te stellen. In plaats daarvan kijken zij veel meer naar de houdbaarheid ervan in het licht van het bewijs dat bestaat.¹²⁵

¹¹⁷ Kuhn 1962, p. 77.

¹¹⁸ Kuhn 1962, p. 77-79.

¹¹⁹ Kuhn 1962, p. 62.

¹²⁰ Kuhn 1962, p. 151.

¹²¹ Kuhn 1962, p. 168.

¹²² Kuhn 1962, p. 77-78.

¹²³ Kuhn 1962, p. 145.

¹²⁴ Kuhn 1962, p. 146-146.

¹²⁵ Kuhn 1962, p. 144.

Toch is ook de wetenschapsopvatting van Kuhn niet zonder kritiek gegaan.¹²⁶ Zo zijn vraagtekens geplaatst bij de interpretatie van de voorbeelden die Kuhn gebruikt. Daaruit kan vaak moeilijk worden opgemaakt of sprake is van normale wetenschap of van een verandering van paradigma. Dat onderscheid blijkt diffuus. De voornaamste kritiek betreft echter het irrationele van Kuhns gedachte, meer in het bijzonder het ontbreken van een neutraal beoordelingskader aan de hand waarvan kan worden bepaald of een paradigma tot vooruitgang heeft geleid. Dit bemoeilijkt het meten van wetenschappelijke vooruitgang. In antwoord daarop heeft Kuhn gesteld dat er wel degelijk waarden zijn waardoor wetenschappelijk onderzoek zich kenmerkt (onder meer: probleemoplossend, accuraat en precies, consistent, reikwijdte, eenvoud en productiviteit), maar dat deze waarden vaag zijn en daarom mogelijk multi-interpretabel. Voorts zijn theorieën (paradigma's) volgens Kuhn aanvankelijk vaak nog onderontwikkeld en verdienen zij verdere doordenking, wat een inschatting van de wetenschappelijkheid op dat moment bemoeilijkt. Dit neemt volgens Kuhn niet weg dat wetenschappers het rationeel oneens kunnen zijn over welke theorie navolging verdient.

Tot slot is er kritiek op de interne consistentie van de gedachtegang van Kuhn. De centrale plaats die Kuhn toekent aan afwijkingen binnen een paradigma vereist dat de afwijking binnen zowel het oude als het nieuwe paradigma als zodanig wordt herkend. Het is niet zeker dat dit gebeurt, want beide paradigma's gebruiken een eigen begrippenkader en een eigen structuur. Het kan dus zo zijn dat de afwijking binnen het nieuwe paradigma niet als zodanig wordt herkend.

5 Lakatos

De kritiek op Kuhn geldt grotendeels ook voor de onderzoeksprogramma's van Imre Lakatos.¹²⁷ Lakatos heeft geprobeerd Poppers falsificatiegedachte en Kuhns paradigmatheorie met elkaar in overeenstemming te brengen. Hij onderkende het belang van het omgaan met afwijkingen van de theorie, maar met Kuhn (en overigens ook Popper) constateerde hij dat in de praktijk niet iedere afwijking onmiddellijk leidt tot het verwerpen van een theorie. Wat Kuhn als paradigma beschouwde, definieerde Lakatos min of meer als een onderzoeksprogramma (*research program*).¹²⁸ Hij beschouwde deze programma's als progressief indien zij zowel een afwijking konden verklaren als nieuwe en succesvolle voorspellingen boden of, met andere woorden, andere, nieuwe afwijkingen

¹²⁶ Zie bijvoorbeeld Laudan 1984, p. 16-18.

¹²⁷ Zie bijvoorbeeld Laudan 1984, p. 20.

¹²⁸ Voor een toelichting van dit begrip, zie Lakatos 1978.

produceerden.¹²⁹ Als aan een van deze voorwaarden niet werd voldaan, betekende dit een stilstand of achteruitgang van het onderzoeksprogramma.

Het spreekt overigens niet voor zich dat het nieuwe paradigma het wint van het oude. Het oude paradigma kan zegevieren, bijvoorbeeld omdat het nieuwe paradigma inzichten biedt die ook het oude paradigma nieuwe impulsen geeft. De wet van de remmende voorsprong wrekt zich dan. In latere handboeken zal de revolutie dan bestempeld worden als een nieuw inzicht dat is ontstaan en een doorbraak heeft betekend binnen het bestaande (oude) paradigma. Maar als het nieuwe paradigma standhoudt, zullen de handboeken het nieuwe paradigma beschrijven als een progressieve beweging die heeft afgerekend met oude, onjuiste theorieën. De handboeken worden immers geschreven door de winnaars van de revolutie.

6 Multicriteria-benadering

Ook op de theorie van Lakatos kwam kritiek, bijvoorbeeld van Paul Feyerabend. Hij stelde zich de vraag waarin de methodologie van een onderzoeker die Lakatos aanhangt zou afwijken van iemand die wars is van methodologische spelregels, en kwam tot de conclusie dat er geen verschil hoefde te bestaan.¹³⁰ Het probleem dat Feyerabend blootlegde, was dat de theorie van Lakatos (en die van Kuhn) louter een beschrijving geeft van hoe theorieën hebben overleefd, maar geen richting geeft aan onderzoekers ten aanzien van wat goed (wetenschappelijk) onderzoek is.¹³¹ Feyerabend beschouwde ‘anything goes’ als de enige methodologische regel in de theorie van Lakatos (en Kuhn) aan de hand waarvan wetenschappelijkheid en vooruitgang in de wetenschap kon worden verklaard.¹³² Wetenschap en de ontwikkelingen en vooruitgang daarin waren te complex om er bruikbare algemene methodologische regels uit te distilleren.¹³³

Feyerabend, en sterker, Larry Laudan, hebben voorts het onderscheid tussen wetenschappelijk en niet-wetenschappelijk, en het onderscheid tussen wetenschap en pseudowetenschap ter discussie gesteld dan wel als diffuus aange-

¹²⁹ Vergelijk de eisen die sommigen thans stellen wil de ene theorie het van de ander kunnen winnen, zie bijvoorbeeld Singleton & Straits 2005, p. 20.

¹³⁰ Feyerabend 1975, met name op p. 186-187. Zie voorts p. 175, waar hij de *epistemological anarchist* introduceert, daarmee het perspectief aangevende waarmee hij Lakatos’ opvatting analyseert.

¹³¹ Feyerabend 1975, p. 185.

¹³² Feyerabend 1975, p. 296.

¹³³ Feyerabend 1975, p. 295 e.v. Zie voorts Bunge 1982, p. 372. Het ‘anything goes’ en ‘against method’ van Feyerabend moet overigens niet worden opgevat als kritiek op het nut van iedere vorm van methodologie, maar vooral als kritiek op eerdere voorstellen en als bewijs dat alle methoden, zelfs de meest voor de hand liggende, beperkingen kennen, zie p. 32.

merkt.¹³⁴ Verschillende auteurs zijn nog op zoek gegaan naar criteria of gezichtspunten aan de hand waarvan wetenschap van pseudowetenschap kan worden onderscheiden, maar de belangstelling lijkt te zijn afgenomen in vergelijking met de jaren waarin Popper, Kuhn, Lakatos, Feyerabend, Laudan, et cetera discussieerden over de wetenschappelijkheid van onderzoek. Dat de vraag naar de wetenschappelijkheid gaandeweg naar de achtergrond is verdrongen, is mogelijk te verklaren door een afgenomen behoefte en noodzaak van wetenschappers om hun werkwijze te verantwoorden. Wetenschappelijk onderzoek en de manier waarop het dient te worden uitgevoerd, staat niet of nauwelijks ter discussie in vergelijking met bijvoorbeeld de 17^e eeuw, toen wetenschap concurrentie had van onder andere de kerk en bovendien de gehanteerde werkwijze in wetenschappelijke studies nog niet algemeen aanvaard was.¹³⁵

De aandacht is, in plaats van naar de wetenschappelijkheid van onderzoek, voor een belangrijk deel uitgegaan naar epistemologische vragen over wat kennis is, hoe het kan worden achterhaald en wanneer het relevant is. Degenen die zich zijn blijven richten op het onderscheid wetenschappelijk/niet-wetenschappelijk, hebben een multicriteria-aanpak ontwikkeld, die bestaat uit een hoeveelheid aan indicatoren aan de hand waarvan kan worden vastgesteld of sprake is van wetenschap of van pseudowetenschap.¹³⁶

¹³⁴ Zie met name Laudan 1983, p. 111-125. Zie voorts Feyerabend 1975, bijvoorbeeld op p. 296 e.v., waar wordt gewezen op overeenkomsten tussen mythes en wetenschap, alsook op de waarde van niet-wetenschappelijke ontdekkingen en praktijken en op de invloed hiervan op wetenschappelijk onderzoek.

¹³⁵ Gutting 2000, p. 431.

¹³⁶ Pseudowetenschap wordt onderscheiden van onwetenschappelijk en niet-wetenschappelijk, hoewel discussie bestaat over de criteria waar aan dient te zijn voldaan wil sprake zijn van een van de drie termen. Overeenstemming bestaat echter over het feit dat sprake moet zijn van niet-wetenschappelijkheid om van pseudowetenschap te kunnen spreken. Om die reden veronderstel ik dat de criteria die zijn geformuleerd om van pseudowetenschap te kunnen spreken, ook gelden voor niet-wetenschappelijk onderzoek. Onwetenschappelijk wordt gelijkgesteld met slechte wetenschap door onderzoekers die tevergeefs wetenschappelijk proberen te zijn. Zie voor het voorgaande, met verdere verwijzingen, Hansson, S.O. (2003). 'Science and Pseudo-Science.' The Stanford Encyclopedia of Philosophy Winter 2003 Edition. Retrieved 17 March 2011, from <http://plato.stanford.edu/entries/pseudo-science/>.

Uit de indicatoren voor pseudowetenschap, die overigens niet voor alle disciplines en ook niet in dezelfde mate hoeven te gelden, valt af te leiden dat ‘[s]cience is a systematic search for knowledge whose validity does not depend on the particular individual but is open for anyone to check or rediscover’.¹³⁷

- *Geloof in autoriteit.* Gesteld wordt dat een bepaald persoon of bepaalde personen over bepaalde eigenschappen beschikken die het mogelijk maken om te bepalen wat juist en wat onjuist is, en dat anderen hun opvattingen of beslissingen dienen te respecteren.
- *Gebrek aan repliceerbaarheid.* Experimenten die niet kunnen worden herhaald door anderen worden toch als betrouwbaar aangemerkt.
- *Selectieve voorbeelden.* Er worden voorbeelden gekozen die niet representatief zijn voor de ‘grote groep’ waar het onderzoek zich op richt.
- *Onwillig om te toetsen.* Een theorie wordt niet getoetst, ook al is dat wel mogelijk.
- *Negeren van informatie.* Observaties of experimenten die conflicteren met de voorgestelde theorie worden genegeerd.
- *Verlaten van verklaringen zonder alternatief voorstel.* Verdedigbare verklaringen worden opgegeven zonder deze te vervangen door een nieuwe verklaring. Hierdoor laat de nieuwe theorie meer onverklaard dan de eerdere theorie.

Hoewel er een redelijke mate van overeenstemming lijkt te bestaan over deze indicatoren, is de herkomst ervan niet altijd even overtuigend. De ‘symptomen’¹³⁸ van onbetrouwbaar wetenschappelijk onderzoek zijn opgesteld aan de hand van een beperkt aantal voorbeelden, op basis van ervaring met vaak een bepaald (beperkt) type publicaties, of zij worden afgeleid uit een publicatie die primair een ander doel heeft dan het onderscheiden van wetenschappelijk onderzoek van niet- of pseudowetenschappelijk onderzoek.¹³⁹ Empirisch onderzoek naar de wetenschappelijkheid van publicaties ontbreekt, voor zover ik heb kunnen overzien.

¹³⁷ Het citaat is ontleend aan de bron in voorgaande noot. De indicatoren die hierna volgen, zijn een vertaling van de indicatoren zoals opgesomd in Hansson, S.O. (2003). ‘Science and Pseudo-Science.’ The Stanford Encyclopedia of Philosophy Winter 2003 Edition. Retrieved 17 March 2011, from <http://plato.stanford.edu/entries/pseudo-science/>, onder verwijzing naar Langmuir 1989 (1953); Gruenberger 1964; Dutch 1982; Radner & Radner 1982; Kitcher 1982; Hansson 1983; Grove 1985; Thagard 1988; Glymour & Stalker 1990; Derksen 1993; Derksen 2003; Vollmer 1993; Ruse 1996; Mahner 2007.

¹³⁸ Doelend op de ‘symptomen van Langmuir’, zie Langmuir 1989 (1953).

¹³⁹ Zie bijvoorbeeld Kitcher 1982 en Ruse 1996, welke publicaties vooral ingaan op creationisme en evolutietheorie.

7 Tussenstand: geen criteria, wel gemeenschappelijke kenmerken

De inductieve benadering van Popper is van groot belang geweest voor de discussie over het onderscheid tussen wetenschappelijk en niet-wetenschappelijk, maar is niet langer dominant binnen de wetenschapsfilosofie en epistemologie. Als de discussie binnen de wetenschapsfilosofie iets heeft opgeleverd, is het wel dat wetenschappelijkheid geen gegeven is en zeker niet iets is dat aan de hand van een algemeen, eenduidig criterium kan worden beoordeeld. Degenen die dat wel menen, waaronder zij die stellen dat het natuurwetenschappelijke model leidend dient te zijn voor andere (wetenschaps)disciplines, lopen het risico het verwijt te krijgen dat zij zich schuldig maken aan *scientism*, een vorm van disciplinair imperialisme waarbij een bepaalde vorm van wetenschap als maatstaf geldt voor, of wordt opgedrongen aan, andere disciplines.¹⁴⁰

Het demarcatiecriterium van Popper is voor anderen aanleiding geweest voor meer inductieve benaderingen van onderzoek. De aandacht is daarbij uitgegaan naar verklaringen voor wetenschappelijke vooruitgang en naar epistemologische vragen over wat (wetenschappelijke) kennis is en hoe het wordt verkregen. Waar de belangstelling uit is gegaan naar het onderscheid tussen wetenschap, niet-wetenschap en pseudowetenschap, heeft dit geresulteerd in auteurs die meerdere indicatoren hebben opgesteld aan de hand waarvan kan worden vermoed of sprake is van pseudowetenschap.

Ook al is er discussie (geweest) over de vraag naar wat wetenschappelijk onderzoek is, lijkt er een zekere mate van consensus te bestaan over een aantal kenmerken van wetenschappelijkheid. In ieder geval keren deze kenmerken geregeld terug in verschillende discussies en opvattingen. Dit hoofdstuk vervolgt met een verkenning van hoe deze kenmerken terugkomen in annotaties en met een verkenning van de gevolgen van het gebruik van criteria van 'wetenschappelijkheid' voor de beoordeling van de kwaliteit van annotaties.

8 Repliceerbaarheid/herhaalbaarheid

Repliceerbaarheid en herhaalbaarheid gaan over de vraag of dezelfde resultaten worden verkregen indien het onderzoek opnieuw of door een ander zou worden uitgevoerd. Zij zijn terugkerende elementen in het debat rondom de wetenschappelijkheid van onderzoek, zowel bij de multicriteria-benadering als bij het falsificatiecriterium van Popper. Om een indruk te krijgen van hoe het zit met de repliceerbaarheid en herhaalbaarheid in annotaties, wordt hierna onderscheid gemaakt tussen aan de ene kant annotaties die regelmatigheden of patronen beschrijven of verklaren en aan de andere kant annotaties die meer evaluatief

¹⁴⁰ Zie Peterson 2003 voor meer informatie over het begrip 'scientism'.

(normatief) van aard zijn. Het zal blijken dat de aard van annotaties bepalend kan (maar niet noodzakelijk hoeft te) zijn voor de wetenschappelijkheid van annotaties.

8.1 *Beschrijven of verklaren van regelmatigheden of patronen*

Een bekend voorbeeld van het beschrijven of verklaren van regelmatigheden of patronen is de situatie waarin de annotator uit een reeks van uitspraken criteria of gezichtspunten distilleert die de genomen beslissing verklaren en die eventueel inzicht geven in hoe het recht moet worden toegepast of hoe het moet worden geïnterpreteerd. De annotatie gaat daarbij verder dan louter een vergelijking in uitkomst. Niet alleen constateert de annotator dat een reeks uitspraken consistent is met dan wel afwijkt van eerdere uitspraken, maar hij verklaart de overeenkomsten en afwijkingen door op zoek te gaan naar criteria of gezichtspunten.

Bijvoorbeeld Gionfriddo heeft op treffende wijze in kaart gebracht hoe juridische onderzoekers te werk gaan als zij een hoeveelheid gerechtelijke uitspraken analyseren en daaruit aantal criteria of gezichtspunten distilleert dat verklaart waarom rechters tot die beslissing komen.¹⁴¹ Aan de hand van een uitgewerkt voorbeeld beschrijft zij het proces van het opstellen, toetsen en bijstellen van hypothesen. De methode van gevalsvergelijking zoals Gionfriddo die beschrijft, is vergelijkbaar met de werkwijze van hoe annotatoren te werk (zouden moeten) gaan als zij gevallen met elkaar vergelijken. Ik vat haar analyse samen.

Het voorbeeld dat Gionfriddo gebruikt, is de vraag wanneer speelgoed als gevaarlijk voor jonge kinderen moet worden aangemerkt. In dit verband heeft zich – fictief – een zevental zaken voorgedaan voor de rechter, te weten de:

- Frisbeezaak: de rechter beoordeelt het gevaar door primair te kijken naar het materiaal waar de frisbee van is gemaakt. Dit leidt naar het oordeel dat een lichte frisbee gemaakt van zacht plastic niet gevaarlijk is.
- Pluizige-popzaak: onder verwijzing naar de frisbeezaak oordeelt de rechter dat een pluizige pop niet gevaarlijk is voor kinderen.
- Speelgoedzwaardzaak: ook hier wordt het speelgoed niet als gevaarlijk aangemerkt. Het speelgoedzwaard is licht en gemaakt van stevig rubber met een rond handvat en een punt aan het einde van het zwaard. De rechter verwijst in zijn beslissing naar de eerdere zaken.
- Honkbalknuppelzaak: een zware, aluminium honkbalknuppel wordt door de rechter wel als gevaarlijk aangemerkt, wederom onder verwijzing naar de eerdere zaken.

¹⁴¹ Gionfriddo 2007. Voor meer informatie over hoe juristen te werk gaan als zij algemene regels afleiden uit één of meer uitspraken, zie Calleros 2001; Smith 1998-1999. Voor een omschrijving van *case synthesis*, zie Greenstein 1985-1986, p. 5.

- Blokkenzaak (I): speelgoedblokken die in elkaar passen zijn gevaarlijk. De scherpe punten van de hoeken kunnen schade veroorzaken aan kinderen, ook al is het speelgoed licht en gemaakt van plastic. De rechter verwijst in zijn uitspraak naar de eerdere uitspraken.
- Balzaak: hier oordeelt de rechter dat een bal, gemaakt van licht maar hard metaal, niet gevaarlijk is voor kinderen. Metaal is, volgens de rechter, niet altijd gevaarlijk voor speelgoed voor jonge kinderen. De rechter verwijst daarbij naar de frisbeezaak, de pluizige-popzaak en de speelgoedzwaardzaak, maar niet naar de honkbalknuppelzaak.
- Blokkenzaak (II). In deze zaak bepaalt de rechter dat ovale blokken gemaakt van stof en gevuld met pluisjes, niet gevaarlijk zijn voor jonge kinderen. De rechter verwijst in dit verband zonder verdere motivering naar de frisbee- en de balzaak.

De eerste stap om een algemene regel op te stellen om te bepalen of speelgoed als gevaarlijk voor jonge kinderen moet worden aangemerkt, is door te kijken naar wat de rechter zelf expliciet oordeelt, aldus Gionfriddo. Wat dan opvalt, is dat de rechter steeds nagaat waar het materiaal van is gemaakt. Dit ondersteunt de hypothese dat het materiaal waarvan het speelgoed is gemaakt (of: hoe het is samengesteld), een criterium is om te beoordelen of het speelgoed gevaarlijk is. Zes van de zeven zaken ondersteunen deze hypothese. In de blokkenzaak (II) gaat de rechter weliswaar niet expliciet in op het materiaal, maar stelt hij wel vast dat het speelgoed is gemaakt van licht en zacht materiaal en verwijst bovendien naar de frisbeezaak en de balzaak.

Een mogelijke vervolgstap is om te veronderstellen dat het materiaal of de samenstelling van het speelgoed het eerste criterium is om te bepalen in hoeverre het speelgoed gevaarlijk is. In de frisbeezaak en in de pluizige-popzaak begint de rechter zijn evaluatie expliciet met het evalueren van het materiaal waarmee het speelgoed is gemaakt. In de andere zaken blijkt dit niet expliciet, maar kan dit worden afgeleid uit de omstandigheid dat de rechter steeds begint met het evalueren van het materiaal. Ook deze hypothese kan dan ook worden bevestigd.

Toch kan het materiaal op zichzelf kennelijk niet verklaren waarom het ene speelgoed als gevaarlijk wordt aangemerkt, en het andere speelgoed niet. Zowel de bal als de honkbalknuppel bestaat uit metaal, maar alleen in de honkbalknuppelzaak wordt een gevaar voor jonge kinderen aangenomen. Kennelijk zijn andere, minder expliciet genoemde criteria relevant naast het materiaal of de samenstelling ervan.

In een aantal uitspraken wordt gerefereerd aan hoe hard of zacht het materiaal is. Dit voert naar de hypothese dat de aard van het materiaal een subcriterium is van het criterium ‘materiaal of samenstelling’ (Tabel 12). De redenering

hierachter is dat zwaarder speelgoed van een bepaald materiaal (bijvoorbeeld metaal) gevaarlijker is dan lichter speelgoed, in die zin dat de kans op letsel en andere schade groter is.

Tabel 12 – Gevalsvergelijking (Gionfriddo)

Materiaal / Samenstelling	Aard van het materiaal	Uitkomst	Zaak
Plastic	Zacht, licht	Ongevaarlijk	Frisbee
Stof en vulling	Zacht	Ongevaarlijk	Pluizige pop
Rubber	Licht, stevig	Ongevaarlijk	Speelgoedzwaard
Plastic	Licht, onbuigzaam	Ongevaarlijk	Blokken (I)
Metaal	Licht, hard	Ongevaarlijk	Bal
Stof en vulling	Zacht	Ongevaarlijk	Blokken (II)
Metaal (aluminium)	Zwaar, onbuigzaam	Gevaarlijk	Honkbalknuppel

Bron: Gionfriddo 2007, p. 24 & p. 26

Op dit punt kunnen de bestaande zeven uitspraken worden verklaard aan de hand van het materiaal of de samenstelling en aan de hand van de aard van het materiaal (hard of zacht). Toch zijn waarschijnlijk nog andere factoren in het spel. De rechter had in de frisbeezaak immers gesteld dat *primair* moest worden gekeken naar het materiaal waar de frisbee van is gemaakt. Daarbij komt dat de rechter in de blokkenzaak (I) ook had meegewogen dat het speelgoed scherpe punten had.

Deze hypothese, dat de vorm ook als een relevant criterium wordt gebruikt om het gevaar van speelgoed te bepalen, stuit op een bezwaar. In de blokkenzaak (I) betreft de rechter de vorm van het speelgoed in zijn beslissing en motivering, maar wordt niet aangenomen dat het speelgoed gevaarlijk is (Tabel 13).

Tabel 13 – Gevalsvergelijking (Gionfriddo)

Materiaal / Samenstelling	Aard van het materiaal	Vorm	Uitkomst	Zaak
Plastic	Zacht, licht	Rond	Ongevaarlijk	Frisbee
Stof en vulling	Zacht	Geen hoeken	Ongevaarlijk	Pluizige pop
Rubber	Licht, stevig	Ronde zijkanten	Ongevaarlijk	Speelgoedzwaard
Plastic	Licht, on-	Scherpe hoeken	Ongevaarlijk	Blokken (I)

	buigzaam			
Metaal	Licht, hard	Rond	Ongevaarlijk	Bal
Stof en vulling	Zacht	Ovaal	Ongevaarlijk	Blokken (II)
Metaal (aluminium)	Zwaar, on- buigzaam	Rond	Gevaarlijk	Honkbalknuppel

Bron: Gionfriddo 2007, p. 31

Op basis van deze zeven uitspraken zijn twee conclusies denkbaar. De eerste is dat de vorm niet als relevant criterium wordt beschouwd in de rechtspraak. In dat geval kan de hypothese worden verworpen. Een tweede conclusie is dat de rechter eerst het materiaal of de samenstelling bekijkt en beoordeelt, en daarna de aard van het materiaal in relatie tot de vorm beoordeelt. Bij deze conclusie hoeft de hypothese niet te worden verworpen. Op basis van de zeven uitspraken kan echter een definitieve beslissing over het al dan niet verwerpen van de hypothese niet worden genomen. De uitspraken geven slechts aanleiding voor twee nieuwe, tegengestelde hypothesen. Gionfriddo concludeert iets vergelijkbaars voor het criterium ‘gebruik’. Het gebruik (onvoorzichtig of niet) kan mede bepalen of speelgoed gevaarlijk is, maar op grond van de zeven uitspraken kan deze hypothese niet worden getoetst.

De analyse van Gionfriddo laat zien hoe de methode van gevalsvergelijking wordt toegepast en hoe deze toepassing kan leiden tot (verdere) theorievorming over wat relevante criteria zijn om de onrechtmatigheid van speelgoed voor jonge kinderen te beoordelen. Het proces dat Gionfriddo beschrijft, illustreert dat het distilleren van criteria of gezichtspunten uit een reeks van uitspraken controleerbaar is en, desgewenst, te repliceren. Het is denkbaar dat een nieuwe analyse daarbij soms tot een andere conclusie kan leiden, bijvoorbeeld omdat de gewezen uitspraak op een verkeerde of andere manier is geïnterpreteerd of opgevat. Dit staat echter los van de repliceerbaarheid of herhaalbaarheid, zo lang duidelijk is welke stappen de annotator heeft gezet en of deze stappen controleerbaar zijn. In welke van de onderzochte annotaties daar sprake van is, is moeilijk vast te stellen, aangezien het een behoorlijke mate van expertise vereist.

Problemen van repliceerbaarheid en herhaalbaarheid zijn dan ook niet aan de orde bij dit type annotaties, althans hoeven dit niet te zijn in het geval de regels van de kunst zijn gevolgd.¹⁴² Het theoretische uitgangspunt (bij welke leer sluit de uitspraak het beste aan, welke gezichtspunten zijn uit de uitspraak te

¹⁴² Van Rhee 2004 en Smits 2009, p. 31 lijken hetzelfde te constateren.

distilleren, ...) is in de regel duidelijk. Indien ook de analyse adequaat is uitgevoerd, maakt dit de annotatie controleerbaar.

8.2 *Evaluatieve (normatieve) annotaties*

Het toetsen aan ‘hoger’ recht is een voorbeeld van het evalueren van een beslissing. Bij het toetsen aan ‘hoger’ recht toetst de annotator of een beslissing in overeenstemming is met bijvoorbeeld een verdragsbepaling of een rechtsbeginsel. De annotator kan nagaan of de theoretische uitgangspunten die ten grondslag liggen aan een bepaald rechtsgebied of bepaalde problematiek op juiste wijze worden toegepast in een concreet geval. Ook kan de annotator verschillen blootleggen tussen de beslissing en wat de theorie voorschrijft, en eventueel suggesties doen over hoe om te gaan met dergelijke verschillen.

Verder kan worden gedacht aan de situatie waarin een beslissing noopt tot een verfijning of aanpassing van de theorie, bijvoorbeeld indien er meerdere theoretische uitgangspunten (beginselen of bepalingen) zijn die met elkaar blijken te botsen. In dergelijke gevallen dringt zich de vraag op of de botsende uitgangspunten met elkaar te verenigen zijn, wat de annotator kan nagaan door de consequenties van verschillende oplossingen naast elkaar te leggen, (eventueel) aan te geven welke oplossing de voorkeur verdient of door te evalueren of de rechter hierin een juiste afweging heeft gemaakt.

Behalve het toetsen aan ‘hoger’ recht, is ook bij het schetsen van implicaties sprake van het evalueren van een uitspraak. Indien de annotator aan de hand van andere, meestal nieuwe casus nagaat wat de gevolgen van de uitspraak zijn en of deze gevolgen wenselijk zijn, evalueert de annotator daarmee de theoretische uitgangspunten die ten grondslag liggen aan de beslissing. Zo kan het schetsen van de gevolgen van een uitspraak resulteren in een conclusie waarin de annotator laat zien dat de uitspraak tot uitkomsten kan of zal leiden die zich slecht verdragen met bepaalde concepten of noties van rechtvaardigheid. De annotator kan in dat verband een of meer oplossingen bespreken die beter voldoen aan die noties, bijvoorbeeld eisen van consistentie of rechtszekerheid.

Een voorbeeld dat dit schetsen van implicaties illustreert, is privaatrechtelijke annotatie nr. 2, die in het eerdere deel van deze studie door drie van de drie deskundigen als goed/uitstekend werd aangemerkt. Deze noot heeft betrekking op een faillissementsrechtelijk leerstuk. De annotator toetst de uitspraak door te bezien of deze leidt tot consistente uitkomsten die redelijk zijn, concludeert dat dit niet het geval is en baseert dit op, kort gezegd, de redenering die de Hoge Raad geeft, zich niet verhoudt tot het doel van de regel, namelijk dat iemand die onverschuldigd heeft betaald aan de gefailleerde niet de dupe mag worden van een vergissing.

Aan deze gedachte liggen bepaalde theoretische uitgangspunten ten grondslag over wanneer het redelijk is om schuldeisers een vordering toe te kennen die voorrang heeft boven alle andere vorderingen en wanneer dit niet redelijk is. Wat die theoretische grondslag is, maakt de annotator niet duidelijk, maar denkbaar is dat de redelijkheid is terug te voeren op uitgangspunten in het faillissementsrecht, waaronder: schulden die voor het faillissement zijn ontstaan, staan lager in de rangorde dan schulden die tijdens faillissement zijn ontstaan, een verplichting van de curator heeft een hogere rang dan een verplichting van de gefailleerde. Auteurs die het standpunt van de annotator delen, doen dit waarschijnlijk vanuit dezelfde theoretische grondslagen als van waaruit de annotator vertrekt. Wie het oneens is met het standpunt, heeft vermoedelijk een ander theoretisch kader over wat redelijk is.

Behalve toetsen of een regel beantwoordt aan een bepaald theoretisch concept door nieuwe gevallen of situaties te bedenken, zoals in de voorbeelden hiervoor, is het ook mogelijk dat de annotator de gewezen rechtsregel of het gebruikte criterium afzet tegen mogelijke alternatieven. De annotator bedenkt dan één of meerdere andere regels, criteria of gezichtspunten die de rechter had kunnen gebruiken, of ontleent deze aan de bestaande literatuur. Vervolgens gaat hij of zij na of die beter invulling geven aan een bepaald theoretisch concept dan de gewezen rechtsregel. Een voorbeeld hiervan is een annotatie waarin de problematiek centraal staat rond de schending van publiekrechtelijke normen die via het privaatrecht worden gehandhaafd. Te denken valt aan de situatie waarin een verzekeraar aan artsen een bonus aanbiedt indien deze bepaalde middelen voorschrijft.¹⁴³ Dergelijke financiële prikkels staan op gespannen voet met Europese regelgeving, die strekt tot bescherming van de volksgezondheid en de prijs waarvoor die wordt aangeboden, maar er niet toe dient om concurrentiebelangen te beschermen. Indien farmaceuten schadevergoeding vorderen van de verzekeraar omdat laatstgenoemde in strijd met Europese regelgeving heeft gehandeld, kunnen deze claims dan ook niet worden toegewezen, aldus de Hoge Raad. De claims stuiten af op het relativiteitsvereiste van art. 6:163 BW.

Een annotator die deze uitspraak bespreekt, zou alternatieven kunnen schetsen. Een alternatief kan zijn dat de rechter in dit geval wel een dergelijke claim had toegestaan. In dat geval zou de vraag rijzen of niet ook andere situaties zich zouden opdringen waar schadevergoeding in beeld komt. Te denken valt aan het geval van hondenbezitters die civielrechtelijk door buurtbewoners aansprakelijk worden gesteld omdat zij hun honden uitlaten op plaatsen die daar niet voor zijn bedoeld (bijvoorbeeld speelplaatsen), of andere vormen waarin publiekrechtelijke normen zijn geschonden en waarvan de handhaving (deels) via het privaatrecht plaatsvindt. Door de gevolgen te schetsen van een dergelijk

¹⁴³ Het voorbeeld is ontleend aan HR 10 november 2006, *NJ* 2008, 491 (*Astrazaneca/Menzis*).

alternatief en deze te koppelen aan de theoretische veronderstellingen waarop het onderscheid tussen straf- en privaatrecht is gebaseerd,¹⁴⁴ toetst de annotator of de theoretische uitgangspunten zijn toegepast die ten grondslag liggen aan, in dit geval, het privaatrecht en de relatie van deze uitgangspunten tot andere rechtsgebieden, in dit geval het publiekrecht.

Een andere illustratie van het evalueren van een uitspraak is de annotatie waarin de annotator een criterium ontwikkelt en vervolgens toetst of de rechter dit criterium heeft toegepast. Dit was aan de orde in de privaatrechtelijke annotatie nr. 4, waar de annotator in essentie, net zoals bij het toetsen aan ‘hoger’ recht en het schetsen van implicaties, de vraag stelt of de rechter het juiste criterium heeft toegepast. In de bedoelde uitspraak stond de kwalificatie van een recht als kwalitatief recht centraal. Die kwalificatie is van belang, omdat een dergelijk recht automatisch mee overgaat met het goed waaraan het verbonden is, en omdat het absolute werking heeft, dat wil zeggen, tegenover een ieder kan worden ingeroepen. Op grond van deze kenmerken betoogt de annotator dat het voor de buitenwereld duidelijk moet zijn dat het recht bedoeld was om mee over te gaan. Hij vertaalt dit in een criterium dat inhoudt dat een direct, onlosmakelijk verband dient te bestaan tussen het recht en het goed. Voor het meewegen van persoonlijke omstandigheden of bedoelingen, zoals voorgesteld in de literatuur, is daarbij geen plaats. Die zijn immers niet zichtbaar voor de buitenwereld.

De grond waarop hij deze opvatting uit de literatuur verwerpt, is de vereiste duidelijkheid die bij goederenrechtelijke mutaties dient te bestaan, zo blijkt uit de annotatie. Deze grond is gerelateerd aan de theoretische grondslag waarop de annotator zijn criterium baseert, te weten het beginsel van publiciteit. Dit houdt in dat een vermogensverschuiving niet verborgen mag blijven, met andere woorden, zichtbaar moet zijn voor anderen.¹⁴⁵ Dit verklaart bijvoorbeeld waarom bij een overdracht van een vordering (openbare cessie) een mededeling is vereist aan de schuldenaar, en waarom bij een stille cessie, waarbij mededeling niet is vereist, de schuldenaar bevrijdend kan betalen aan de persoon die de vordering stil heeft overgedragen.

Het idee achter het theoretisch concept van het beginsel van publiciteit is dat ‘publicatie’ bijdraagt aan het vertrouwen en, daarmee, aan de stabiliteit van het handelsverkeer. Zou publiciteit ontbreken, dan zou, in het voorbeeld van de cessie, een schuldenaar steeds moeten achterhalen of degene aan wie hij betaalt, ook daadwerkelijk degene is aan wie hij moet betalen. Een afname van het vertrouwen in dergelijke transacties resulteert waarschijnlijk in minder transacties (handel) en in hogere onderzoeks- en transactiekosten (de schuldenaar moet onderzoeken of hij moet betalen aan wie hij denkt te moeten betalen). Deze

¹⁴⁴ Bijvoorbeeld: het strafrecht is beter uitgerust om publiekrechtelijke normen te handhaven dan het privaatrecht, onder meer in verband met de waarborgen die het biedt.

¹⁴⁵ Zie bijvoorbeeld Snijders & Rank-Berenschot 2007, nr. 80.

normatieve theorie lijkt ten grondslag te liggen aan het criterium dat de annotator ontwikkelt en waarvan hij bekijkt of het is overgenomen door de rechter.

Wat vanuit het perspectief van repliceerbaarheid en herhaalbaarheid met name problematisch is bij evaluatieve (normatieve) uitspraken, is dat de theoretische uitgangspunten die ten grondslag liggen aan een bepaalde analyse niet altijd worden geconcretiseerd dan wel impliciet blijven. Dit bemoeilijkt het repliceren van het onderzoek door een buitenstaander. Met andere woorden, repliceerbaarheid hangt sterk af van hoe concreet de gekozen theoretische uitgangspunten zijn.

Een voorbeeld waaruit de moeilijkheid blijkt om een evaluatieve, normatieve analyse te repliceren, is de eerder genoemde privaatrechtelijke annotatie waarin de annotator concludeert dat de uitspraak niet tot uitkomsten leidt die redelijk en consistent zijn. In de uitspraak stelt de annotator vast dat de Hoge Raad een criterium heeft genuanceerd, waarna hij nagaat tot welke uitkomsten die nuancering in andere gevallen leidt en vervolgens concludeert dat de regel inclusief nuancering tot onredelijke uitkomsten leidt. Wat hier vanuit het oogpunt van repliceerbaarheid met name problematisch is, is dat het concept van rechtvaardigheid niet wordt geconcretiseerd. Hierdoor is het moeilijk voor een buitenstaander om na te gaan waarom de regel die de rechter heeft gewezen als onredelijk of onwenselijk moet worden gezien. Ook is het moeilijk te controleren of de redenering die tot die conclusie heeft geleid houdbaar of gebrekkig is. Dit neemt overigens niet weg dat de analyse en de conclusie kan aanspreken, in die zin dat anderen het eens zijn met de opvattingen en conclusies van de annotator. Een opvatting delen valt echter niet samen met het kunnen construeren van de totstandkoming van de opvatting of redenering, of met het kunnen definiëren of achterhalen van de onderliggende theoretische uitgangspunten.

Toch hoeft repliceerbaarheid als gezegd niet altijd problematisch te zijn bij normatieve theorievorming. Het andere voorbeeld dat ik hiervoor gebruikte illustreert dit. Het gaat om de annotatie waarin een criterium is opgesteld aan de hand van het beginsel van publiciteit. Aan de hand van dit beginsel betoogt de annotator dat een recht automatisch mee overgaat met een ander recht indien de rechten aan elkaar zijn verknocht. De omstandigheid dat de annotator het beginsel expliciteert waaraan hij toetst, maakt het voor een buitenstaander mogelijk om na te gaan of het criterium dat de annotator introduceert, daadwerkelijk een adequate vertaling of concretisering vormt van het genoemde beginsel. Het verschil met het vorige voorbeeld is dat de annotator in deze annotatie één beginsel kiest dat bovendien voldoende scherp is omschreven om te toetsen of het gebruikte of ontwikkelde criterium overeenstemt met het gekozen uitgangspunt.

In annotaties waarin wordt getoetst welke leer wordt toegepast, lijkt repliceerbaarheid ook niet problematisch. Dit kenmerk is aan de orde in een tweetal strafrechtelijke annotaties dat eerder in dit onderzoek is geanalyseerd. In deze

annotaties toetste de annotator welke leer de rechter had toegepast. In de literatuur bestonden twee stromingen, die een verschillende leer aanhingen. In de ene annotatie laat de annotator zien dat het toepassen van de andere leer tot een andere beslissing zou hebben geleid dan in de uitspraak in kwestie te lezen valt. In de andere annotatie waar de annotator toetst welke leer is toegepast, betoogt de annotator dat de uitspraak alleen goed te begrijpen valt vanuit de ene leer, maar niet vanuit de andere leer.

Repliceerbaarheid lijkt dan ook vooral een probleem bij annotaties waar de theoretische uitgangspunten die ten grondslag liggen aan bepaalde conclusies niet worden geëxpliciteerd, indien deze uitgangspunten algemene, niet-geconcretiseerde beginselen zijn, zoals redelijkheid of rechtvaardigheid, of indien de annotator toetst aan meerdere beginselen waarvan niet duidelijk is hoe zij zich tot elkaar verhouden of welk beginsel in welke situatie voorgaat op het (de) andere beginsel(en).

9 Andere kenmerken van pseudowetenschappelijkheid en hun relatie tot annotaties

Net als voor repliceerbaarheid en herhaalbaarheid geldt voor de andere indicatoren voor pseudowetenschap dat het moeilijk is om in zijn algemeenheid te bepalen in hoeverre annotaties daar aan voldoen. De inhoud van annotaties en de werkwijze van annotatoren verschillen daarvoor te veel. Ook veronderstelt een oordeel over deze kenmerken dat de beoordelaar in hoge mate ingevoerd is in een bepaald vakgebied, maar ook in de werkwijze van de annotator. Ontbreekt die kennis, dan is het moeilijk inschatten of een annotator selectieve voorbeelden gebruikt dan wel bronnen opzettelijk niet noemt om een redenering kracht bij te zetten. In een eerdere publicatie is voor een reeks juridische artikelen voorts vastgesteld dat een aantal ervan goed scoort op punten als betrouwbaarheid, evenwichtigheid, en niet-selectiviteit. In een aantal artikelen waren er meer problemen met de betrouwbaarheid, evenwichtigheid en niet-selectiviteit van de gebruikte bronnen.¹⁴⁶

De gevonden kwaliteitsindicatoren geven geen bewijs voor, maar ook niet tegen mogelijke selectiviteit of het negeren van informatie. Wel zijn er, meer in het algemeen, aanwijzingen gevonden die in de richting wijzen dat een zorgvuldige bewijsvoering is van belang om van een goede annotatie te kunnen spreken. De kwantitatieve analyse die eerder in dit onderzoek is uitgevoerd, leverde op dat er een relatie lijkt te bestaan tussen de kwaliteit van annotaties enerzijds en

¹⁴⁶ Van Gestel & Vranken 2007. Maar, zo voegen zij er aan toe, ‘niet uitgesloten is dat dit samenhangt met de in die artikelen gehanteerde stijl en publicatievorm. Was het vooraf wel de bedoeling van de auteurs om in de rubriek Wetenschap te publiceren of hadden zij zelf bijvoorbeeld een essay of opinie voor ogen toen ze hun bijdrage schreven?’

het gebruik van de bronnen jurisprudentie (beoordelaar C) en literatuur (beoordelaar A), alsook een gebrek aan bronnen (beoordelaar C) anderzijds.¹⁴⁷

Uitvoeriger over de bewijsvoering binnen annotaties is de eerder aangehaalde studie van Crombag & Cohen. Zij hebben een eerste aanzet gedaan om in kaart te brengen hoe een annotator te werk gaat bij het schrijven van een annotatie. De instructie die Crombag & Cohen aan de annotator in kwestie (A.R. Bloembergen) meegaven, was om een taperecorder ‘bij de hand te houden zolang hij met het arrest bezig was’ en om ‘zijn denkwerk zo precies mogelijk vast te leggen op de band’.¹⁴⁸ Dat leverde tientallen pagina’s op aan uitgewerkt verslag. Ik beperk mij tot een samenvatting ervan.¹⁴⁹

De annotator begint met het distilleren van het essentiële punt uit de uitspraak (het Fluorideringsarrest). Dat is, volgens de annotator, de verhouding tussen wetgever en rechter. De annotator stelt vast dat de rechter in dit geval geen politieke knoop doorhakt; dat is aan de wetgever, aldus de rechter. Hij herinnert zich echter een geval in het stakingsrecht, waar de rechter wel een politieke knoop had doorgemaakt. Volgens de annotator komt dat door de concrete wettelijke regeling die bij staking niet bestond en die in dit geval wel bestond.

Crombag & Cohen merken op dat de annotator al in het eerste stadium de belangrijkste beslissing heeft genomen voor de noot. Ik voeg daar aan toe dat de annotator hier een hypothese heeft gevormd over het criterium op basis waarvan het verschil in uitkomst tussen het Fluorideringsarrest en het eerdere arrest over stakingsrecht kan worden verklaard.

Nadat de annotator het essentiële punt heeft achterhaald uit de uitspraak, gaat de annotator over tot het nader lezen van het arrest, het zoeken naar relevante literatuur, jurisprudentie en naar gevallen waarin vergelijkbare problematiek aan de orde is. Hij refereert daarbij aan de essentiële passages uit de uitspraak en plaatst vraagtekens bij de overtuigingskracht ervan. De kern van het geschil blijkt te liggen in het oordeel van de Hoge Raad, die een wettelijke bepaling op een manier interpreteert waardoor de kwestie er een wordt waar de rechter zijn handen van af moet houden. Die beslissing had ook de andere kant kunnen uitvallen, aldus de annotator, want uit de wetsgeschiedenis volgt dat de wetgever niet had stilgestaan bij de situatie die zich in dit arrest voordoet, namelijk fluoridering van drinkwater.

Voorts constateert de annotator dat de lijn van de Hoge Raad afwijkt van die van de Kroon. Dit vormt het tweede thema van de annotatie. Het derde thema – mensenrechten en gewetensbezwaren – komt de annotator op het spoor door andere publicaties over het onderwerp te lezen. Tussendoor merkt de annotator op dat het inspreken op de taperecorder invloed heeft op de manier waarop

¹⁴⁷ Hoofdstuk 2, paragraaf. 4.

¹⁴⁸ Crombag & Cohen 1974, p. 2-3.

¹⁴⁹ Crombag & Cohen 1974, p. 5a-44b.

hij een arrest aanpakt. De recorder dwingt tot het ordenen van de gedachten, welke ordening anders minder is.

Vervolgens denkt de annotator vooral na over hoe de informatie te structureren, over de vraag welke aspecten wel of niet aan de orde moeten komen en over hoe veel hij uit moet weiden over bepaalde aspecten. Dit betreft vooral uitwerking. De annotator beschrijft vooral welke bronnen hij erop naslaat en verwerkt, alsook het schrijfproces en de worstelingen die hij daarbij doormaakt.

Wat aan de beschrijving van de bewijsvoering opvalt, is dat de annotator gaandeweg aanwijzingen vindt dat zijn hypothese, dat het bestaan van een wettelijk bepaling beslissend is voor de vraag of een rechter zich uit mag laten over politiek geladen vraagstukken, bevestigt.¹⁵⁰ Tegelijk wekt de omschrijving van het proces de indruk dat de annotator open staat voor informatie die zijn oorspronkelijke hypothese zou kunnen ontkrachten. Dit is een voorzichtige aanwijzing dat selectiviteit niet onder een zorgvuldige bewijsvoering valt, maar niet kan worden uitgesloten dat het open staan voor informatie die een andere richting op wijst voortvloeit uit sociaalwenselijk gedrag.

Ten aanzien van een andere indicator voor pseudowetenschap, te weten het geloven in een bepaalde autoriteit, is niet gebleken dat dit aan de orde is in annotaties. Uit de eerder uitgevoerde analyse van annotaties volgt niet dat bepaalde personen bepalen wie gelijk en wie ongelijk heeft, of dat wordt gesteld dat anderen de opvattingen of beslissingen van deze personen dienen te respecteren. Wat wel opvalt, is de volgzzaamheid in annotaties.¹⁵¹ Annotaties, ook de goede/uitstekende, stellen de uitspraak als zodanig in beperkte mate ter discussie. Wat in de regel gebeurt, is dat annotatoren analyseren of bekritisieren hoe de uitspraak verhoudt tot eerdere uitspraken of bepaalde beginselen, of wat de gevolgen ervan zijn voor de wetsystematiek en de praktijk. Minder gebruikelijk is dat de annotator de vraag opwerpt wat er zou zijn gebeurd als de rechter een andere beslissing zou hebben genomen en of die beslissing beter zou zijn geweest.

De volgzzaamheid van annotaties brengt mee dat de annotator uitgaat van de uitspraak en hieruit gezichtspunten of criteria distilleert, de uitspraak tracht in te passen in het wettelijke systeem, nagaat of de uitspraak aansluit bij eerdere uitspraken, of de gevolgen ervan bespreekt voor andere leerstukken of rechtsfiguren. Dit suggereert dat een andere indicator van pseudowetenschap, te weten het verlaten van verklaringen zonder een alternatief aan te dragen, niet aan de orde is in annotaties waarin wordt voortgebouwd op bestaande concepten en uitgangspunten. Het geldende recht is immers vaak leidend.

In gevallen waarin de annotator minder volgzzaam is, kan het verlaten van verklaringen zonder het geven van een alternatief wel aan de orde zijn. In een

¹⁵⁰ Zie Crombag & Cohen 1974, p. 30a.

¹⁵¹ Hoofdstuk 2, paragraaf 5.

annotatie die eerder in deze studie is onderzocht, die als goed/uitstekend was aangemerkt en waarin sprake was van het schetsen van implicaties, gebeurde het dat de annotator de uitspraak ‘doorrekende’ naar andere, nieuwe situaties. Naar aanleiding hiervan werd geconcludeerd dat het gebruikte criterium tot onredelijke resultaten leidt. De annotator introduceerde zelf een alternatief, maar dit week nauwelijks af van de door de rechter geformuleerde regel. Bovendien concludeerde de annotator zelf dat het alternatief vermoedelijk niet de onredelijkheid zou wegnemen.

Aanwijzingen dat annotatoren onwillig zouden zijn om theorieën te toetsen zijn, tot slot, niet gevonden in de onderzochte annotaties. Zoals hiervoor beschreven, worden wel degelijk theorieën gevormd en getoetst, zij het dat niet altijd duidelijk is wat het onderliggende theoretische concept is. Dit staat echter los van de bereidwilligheid om te toetsen.

10 Experimentele studie naar effect van aanleggen wetenschappelijke criteria op beoordeling van annotaties

10.1 Onderzoeksvraag en opzet

Om zicht te krijgen op het effect van het gebruik van wetenschappelijke criteria, is nagegaan of het instrueren van deskundige beoordelaars met wetenschappelijke criteria gevolgen heeft voor hun beoordeling van de kwaliteit van annotaties. Daartoe is aan deskundigen gevraagd een score toe te kennen aan een annotatie, waarbij een deel van de respondenten is geïnstrueerd de annotatie te beoordelen aan de hand van een aantal specifieke wetenschappelijke criteria. Het andere deel van de respondenten ontving geen instructie en was vrij om de annotatie te beoordelen. De verwachting was dat de instructie tot gevolg zou hebben dat de beoordelingen tussen de beoordelaars minder uiteen zouden gaan lopen.

H₃: Het instrueren van beoordelaars om annotaties te beoordelen aan de hand van wetenschappelijke criteria leidt tot andere kwaliteitsoordelen dan zonder deze instructie

10.2 Methode

Sample

Een elektronische vragenlijst is verstuurd aan 69 personen. De deelnemers waren gepromoveerde strafrechtjuristen verbonden aan de Nederlandse rechtenfaculteiten. Net als in voorgaande studies was de verwachting dat gepromoveerden in dit rechtsgebied voldoende in staat zouden zijn om de kwaliteit van een annotatie te beoordelen. Gepromoveerden die betrokken zijn geweest bij eerdere

fasen van het onderzoek of anderszins bekend zijn met de inhoud van het onderzoek zijn niet benaderd in verband met mogelijke beïnvloedingseffecten.

Van de 69 personen die zijn aangeschreven, zijn er elf begonnen aan de vragenlijst en hebben er negen de gehele vragenlijst ingevuld. De twee personen die vroegtijdig zijn afgehaakt, hebben dit gedaan na de tweede vraag/stelling. Deze respondenten zijn niet in de analyse betrokken. Dit brengt de respons op 13%. De respondenten zijn gelijk over de condities verdeeld: vijf personen in de groep die geen instructie/criteria kreeg en vier personen in de groep met instructie/criteria. In de eerste groep zijn de respondenten gemiddeld 41 jaar, is het gemiddelde promotiejaar het jaar 2000, zijn vier van de vijf respondenten in 2000 of daarna gepromoveerd en is één van de vijf respondenten van het vrouwelijke geslacht. In de groep met instructie/criteria is de gemiddelde leeftijd op 38 jaar, is het gemiddelde promotiejaar 2005, zijn alle respondenten in 2000 of daarna gepromoveerd en is één van de vijf respondenten van het vrouwelijke geslacht. Alle deelnemers gaven aan gespecialiseerd te zijn in het strafrecht. Een manipulatiecheck bevestigde dat de ene groep geen instructie/criteria had ontvangen en de andere groep wel.

De annotatie die is voorgelegd, is een strafrechtelijke annotatie waar de beoordelaars een verschillende score aan hadden toegekend. Reden hiervoor is dat bij een dergelijke annotatie er een kans bestaat dat ook andere deskundigen verschillende scores aan de annotatie toekennen. Van de vier strafrechtelijke annotaties die in aanmerking kwamen, is annotatie nr. 12 gekozen. Deze annotatie bestaat uit meer kenmerken dan het louter beschrijven of samenvatten van de uitspraak.

Procedure

Aan de deelnemers is verzocht de annotatie door te nemen en hier een score aan toe te kennen (Likertschaal, zevenpuntschaal). Een deel van de deelnemers ontving een instructie over welke criteria zij dienden aan te leggen, inclusief een omschrijving van deze criteria. De andere deelnemers ontving deze instructie en criteria niet. De deelnemers zijn random over de groepen verdeeld.

Degenen die de instructie/criteria hadden ontvangen, dienden de annotatie voor zowel ieder criterium afzonderlijk als voor de criteria gezamenlijk te beoordelen. Op deze manier kon mogelijk worden achterhaald of een bepaald criterium een grotere rol had gespeeld dan een ander criterium. Na de beoordeling kregen de deelnemers een reeks stellingen te beoordelen. Ook deze hadden tot doel te achterhalen welk criterium de doorslag had gegeven. Daarna is gevraagd of zij de annotatie anders zouden hebben beoordeeld als zij de annotatie zonder instructie zouden moeten hebben beoordeeld. Het bieden van deze mogelijkheid had tot doel te voorkomen dat de respondenten, tegen de instructie in, eigen criteria zouden aanleggen om de annotatie te beoordelen. Daarnaast zou de ver-

gelijking tussen de beoordeling met en zonder de instructie/criteria extra inzicht kunnen geven in het effect van het gebruik van wetenschappelijke criteria. Bij de eerste beoordeling was aangekondigd dat deelnemers later de mogelijkheid hadden om de annotatie los van de aangereikte criteria te beoordelen. Vervolgens is aan de deelnemers gevraagd hun beoordeling toe te lichten (open vraag).

Aan degenen die geen instructie/criteria hadden gekregen is gevraagd de annotatie te beoordelen en vervolgens uit te leggen waar zij hun beoordeling op hadden gebaseerd (open vraag). Aan alle respondenten zijn stellingen voorgelegd om te controleren of de manipulatie (wel/geen instructie) was geslaagd en om hun affiniteit met het schrijven van annotaties alsook hun algemene opvattingen hierover te meten. Verder is gevraagd naar demografische kenmerken (geslacht, leeftijd, etc.), alsook naar of zij wisten wie de annotator van de betreffende annotatie was.

10.3 Resultaten

De statistische analyses die zijn uitgevoerd, laten geen verbanden zien, vermoedelijk vanwege het lage aantal respondenten. De weergave van de resultaten beperkt zich daarom tot een weergave van wat is geobserveerd (Tabel 14), zonder te (kunnen) berekenen of de gevonden verschillen waarschijnlijk op toeval berusten.

Tabel 14 – Resultaten

		Zonder instructie / criteria	Met instructie / criteria
Algeheel oordeel	Gemiddelde	5.60	5.50
	Laagste	5	5
	Hoogste	6	6
Bronnengebruik	Gemiddelde	-	5.00
	Laagste	-	4
	Hoogste	-	6
Verdedigbaarheid	Gemiddelde	-	5.50
	Laagste	-	5
	Hoogste	-	6
Alternatieven	Gemiddelde	-	4.75
	Laagste	-	4
	Hoogste	-	6
Oordeel zonder instructie/criteria	Gemiddelde	-	5.50
	Laagste	-	5
	Hoogste	-	6

Noot: scores zijn gemeten met een zevenpunt-Likertschaal.

De drie criteria die de helft van de deelnemers hebben gekregen, zijn bronnengebruik, verdedigbaarheid en het schetsen van alternatieven bij kritiek. Bij bronnengebruik gaat het, in het kort, om de vraag of selectief is omgegaan met informatie en dus of de annotator voldoende oog heeft gehad voor andere standpunten of geluiden. Zo niet, dan scoort de annotatie laag op dit onderdeel.

Het criterium van verdedigbaarheid is een operationalisering van het criterium van repliceerbaarheid. Verdedigbaarheid houdt hier in dat het standpunt in redelijkheid verdedigbaar is. Dit wil niet zeggen dat de beoordelaar het eens is met de annotatie, maar wat de annotator schrijft dient wel verdedigbaar te zijn, in die zin dat een beoordelaar moet kunnen nagaan hoe de annotator tot een bepaalde conclusie is gekomen en, gegeven de (geëxpliciteerde) invalshoek, zelf tot een vergelijkbare conclusie zou komen.

Het criterium van ‘alternatieven’ houdt in dat de annotator alternatieven bespreekt indien deze kritiek uit op de uitspraak in kwestie. Die alternatieven kunnen variëren van het aandragen van criteria tot het formuleren van een andere uitkomst tot het geven van een andere motivering.

Het blijkt dat de algemene oordelen niet of nauwelijks verschillen tussen de twee groepen ($M = 5.60$ versus $M = 5.50$). De spreiding is gering en daarmee identiek tussen de twee groepen: de laagst gegeven score is 5 en de hoogste score 6. Interessant is voorts dat de respondenten die de annotatie beoordeelden aan de hand van de drie hiervoor genoemde criteria, dezelfde score toekennen aan de annotatie ($M = 5.50$) als dat zij in eerste instantie hadden gedaan. Dit geldt voor alle respondenten in die groep. Met andere woorden, de beoordelin-

gen op grond van de wetenschappelijke criteria in de groep met instructie/criteria wijken niet af van beoordelingen van diezelfde beoordelaars als zij vervolgens hun eigen criteria aanleggen. Deze scores komen logischerwijs dan ook opnieuw overeen met de beoordelingen in de andere groep, dat wil zeggen de groep die geen instructie/criteria hadden ontvangen.

Kijkend naar de scores die de respondenten hebben gegeven voor de afzonderlijke criteria, valt op dat er weinig verschillen zijn tussen de criteria onderling: de gemiddelde scores variëren tussen de 4.75 en 5.50, met ook hier weinig spreiding (laagste score = 4, hoogste score = 6). Alleen het criterium over het schetsen van alternatieven springt er enigszins uit met een gemiddelde van 4.75, maar dit wordt veroorzaakt door één relatief lage score die het gemiddelde snel omlaag trekt in een kleine steekproef als deze.

10.4 Discussie

De verwachting was dat het instrueren van beoordelaars om annotaties te beoordelen aan de hand van wetenschappelijke criteria tot gevolg zou hebben dat zij tot een ander kwaliteitsoordeel zouden komen dan als zij deze instructie niet zouden hebben gehad (H_3). Deze verwachting is niet uitgekomen. Op basis van de (weinig) gegevens die konden worden verzameld, moet vooralsnog worden vastgesteld dat het effect van de instructie/criteria verwaarloosbaar is. De groep die de instructie/criteria heeft ontvangen, komt tot dezelfde kwaliteitsoordelen als de groep die de instructie/criteria niet heeft gekregen, in dit geval zelfs tot precies dezelfde scores.

Het aantal deelnemers aan het experiment is te klein om vast te stellen of de gevonden overeenkomsten en verschillen op toeval berusten. Wel is duidelijk dat als het gevonden patroon zich door zou zetten, het effect van de wetenschappelijke criteria (verwaarloosbaar) klein is. Het valt echter niet uit te sluiten dat bij een grotere steekproef de resultaten anders zouden zijn, dat wil zeggen, dat de instructie/criteria toch tot andere beoordelingen zouden leiden. Verder is het mogelijk dat bij een andere annotatie andere resultaten zouden zijn verkregen. Tot slot is het denkbaar dat de deelnemers aan het experiment, in weerwil van de instructie, de instructie/criteria hebben genegeerd en de annotatie toch aan de hand van een eigen toetsingskader hebben beoordeeld.

Niettemin zijn de resultaten, hoe voorlopig ook, interessant. Ze wekken de indruk dat de criteria die het juridisch-academisch forum aanlegt om annotaties te beoordelen, mogelijk niet wezenlijk verschillen van criteria die als wetenschappelijk worden aangemerkt.

11 Conclusie

Het is niet eenvoudig om uitspraken te doen over het wetenschappelijke karakter van annotaties. In de wetenschapsfilosofie is het demarcatiecriterium niet langer onderwerp van discussie. Bovendien wordt onderkend dat de eisen die aan wetenschappelijke publicaties worden gesteld, per discipline verschillen, in ieder geval kunnen verschillen. Daarnaast is het belangrijk voor ogen te houden dat het debat over de wetenschappelijkheid van onderzoek zoals dat in de wetenschapsfilosofie is gevoerd, niet tot doel heeft gehad criteria aan te reiken aan de hand waarvan individuele publicaties kunnen worden beoordeeld. Het lijkt vooral een zoektocht te zijn (geweest) naar beschrijvingen en verklaringen voor het verschijnsel wetenschap.

Het voorgaande neemt niet weg dat de discussie relevante inzichten heeft opgeleverd over een aantal kenmerken van annotaties met betrekking tot de wetenschappelijkheid ervan. Waar weinig discussie over is, is dat wetenschappelijk onderzoek nastreeft replicateerbaar te zijn. Het toepassen van dit principe op annotaties leidt tot de vaststelling dat bij met name evaluatieve (normatieve) annotaties het gevaar op de loer ligt dat uitspraken afhankelijk zijn van de individuele onderzoeker. Repliceerbaarheid is dan in het geding, vooral als de annotatie de onderliggende theoretische uitgangspunten onvoldoende specificeert, bijvoorbeeld in het geval van redelijkheid of rechtvaardigheid, of indien het theoretische concept bestaat uit meerdere beginselen waarvan niet duidelijk is hoe de beginselen zich tot elkaar verhouden en welk beginsel in welke situaties het meeste gewicht krijgt. In dergelijke gevallen is het moeilijk na te gaan hoe de annotator tot een bepaalde conclusie of tot een bepaald standpunt is gekomen. Het moge zo zijn dat dit gebrek aan controleerbaarheid te verklaren is, onvermijdelijk is, of als wenselijk wordt beschouwd, maar een dergelijke insteek is niet bevorderlijk voor de replicateerbaarheid, en daarmee ook niet voor de wetenschappelijkheid.

Bij annotaties die beschrijvend of verklarend van aard zijn, lijkt replicateerbaarheid minder problematisch. Een beschrijving van hoe dit proces in zijn werk gaat, laat zien dat het voor een buitenstaander, in vergelijking met evaluerende (normatieve) annotaties, veel beter mogelijk is om het proces van het opstellen, bijstellen en verfijnen van hypothesen te controleren en desgewenst te repliceren.

Voor andere kenmerken van wetenschappelijkheid (selectieve informatie, negeren van informatie, geen alternatieven ontwikkelen bij het geven van kritiek) is het moeilijk om vast te stellen in hoeverre zij wel of niet terugkomen in annotaties. Om dit goed te beoordelen, is deskundigheid vereist ten aanzien van het betreffende rechtsgebied dan wel het individuele leerstuk.

Wat levert het voorgaande op voor de relatie tussen de eerder in deze studie gevonden kwaliteitsindicatoren en de wetenschappelijkheid van annotaties? De gevonden kwaliteitsindicatoren hebben vooral betrekking op de inhoud van annotaties en minder op de werkwijze en de manier waarop annotaties worden geschreven. Voor de kenmerken van wetenschappelijkheid geldt het omgekeerde: deze zien vooral op de werkwijze (van een discipline of een groep van onderzoekers) maar niet op de inhoud van het onderzoek.

Er zijn ook overeenkomsten. Waar wetenschappelijkheid doorgaans te maken heeft met theorieën en theorievorming, is in deze studie vastgesteld dat annotaties die systematische kennis produceren, dat wil zeggen, kennis die het individuele geval ontstijgt, een grotere kans hebben om als goed of uitstekend te worden aangemerkt dan annotaties die conclusies produceren die op de individuele uitspraak betrekking hebben.

Ook laten de inzichten over wetenschappelijkheid van onderzoek zien er een grotere kans bestaat op problemen met de repliceerbaarheid bij evaluatieve (normatieve) annotaties in vergelijking met annotaties die regelmatigheden of patronen beschrijven of verklaren. Dit inzicht is relevant gezien vanuit het oogpunt van wetenschappelijkheid. In de analyse die eerder in dit onderzoek is uitgevoerd, kwam naar voren dat normatieve diepgang het overkoepelende criterium is aan de hand waarvan de kwaliteit van annotaties wordt beoordeeld.¹⁵² Diezelfde analyse leverde op dat het vergelijken van de becommentarieerde uitspraak met eerder verschenen jurisprudentie over het algemeen bestaat uit constatering (‘de uitspraak komt wel/niet overeen met eerdere uitspraken’), wat dan ook niet tot een hogere kwaliteit van annotaties leidde. Ik heb geen annotaties aangetroffen waarin (relatief) uitgebreid uitspraken met elkaar worden vergeleken om daar gezichtspunten, criteria of anderszins patronen uit af te leiden, laat staan op de manier zoals Gionfriddo die beschrijft. Met uitzondering van een aantal strafrechtelijke annotaties, waarin aan de hand van een of meerdere uitspraken werd onderzocht welke theorie het beste aansloot bij de becommentarieerde uitspra(a)k(en), is het beschrijven of verklaren van regelmatigheden of patronen nauwelijks aan de orde in annotaties, in vergelijking met evaluerende (normatieve) annotaties en afgaand op de 48 aangeleverde publicaties. Annotaties die evaluatief (normatief) van aard zijn, zijn veel gebruikelijker, maar kwetsbaarder op het punt van repliceerbaarheid en dus op het punt van wetenschappelijkheid.

De experimentele studie die is uitgevoerd om een annotatie met en zonder een instructie aan de hand van wetenschappelijke criteria te beoordelen, schetst echter het beeld dat het expliciet opnemen van repliceerbaarheid/verdedigbaarheid als criterium niet tot wezenlijk andere beoordelingen zal

¹⁵² Zie Hoofdstuk 3 voor de grenzen van dit criterium.

leiden. Mogelijk wordt repliceerbaarheid (impliciet) al als criterium gebruikt, maar niet kan worden uitgesloten dat de resultaten het gevolg zijn van het kleine aantal observaties, van een homogene groep van respondenten of van strategische antwoorden van misschien al bevooroordeelde respondenten.

Een verklaring voor het gegeven dat het beschrijven of verklaren van regelmatigheden of patronen niet vaak terugkomt in annotaties, lijkt te zijn dat annotatoren (of juridische onderzoekers) beperkte mogelijkheden hebben om annotaties te schrijven waarin gezichtspunten of criteria uit uitspraken worden afgeleid, althans in Nederland. In de praktijk wordt het gras vaak al voor de voeten van de annotator weggemaaid. Door de motiveringsplicht die geldt voor gerechtelijke uitspraken en de conclusie van de Advocaat-Generaal in het geval van een uitspraak van de Hoge Raad, zal een rechter in zijn uitspraak zelf al redenen, gezichtspunten of criteria hebben genoemd waarop de beslissing is gebaseerd. Dit beperkt de speelruimte van de annotator aanzienlijk, wat verklaart dat annotatoren uitspraken ‘doorrekenen’ naar leerstukken of rechtsgebieden die niet aan de orde zijn in de uitspraak.

Het voorgaande verklaart mogelijk tevens, in ieder geval deels, waarom annotaties in Frankrijk anders (positiever) worden gewaardeerd dan in bijvoorbeeld Nederland. De *Cour de Cassation* motiveert haar uitspraken niet of nauwelijks. In vergelijking met annotatoren in Nederland geeft dit Franse annotatoren, en de juridische gemeenschap in Frankrijk als geheel, meer mogelijkheden om regelmatigheden en patronen in een reeks van uitspraken te beschrijven of te achterhalen. De (veronderstelde) hogere motiveringseisen voor rechters in Nederland kunnen mede het verminderde aanzien van annotaties in termen van academische of wetenschappelijke status verklaren. Of dat daadwerkelijk zo is, is moeilijk te achterhalen.

Naar een beoordeling van de (academische) kwaliteit

1 Inleiding

Aan het slot van deze studie rijst de vraag naar de betekenis van de resultaten en inzichten voor de beoordeling en status van annotaties. Er zijn grofweg drie denkrichtingen te onderscheiden, die elkaar overigens niet uitsluiten: een procedurele richting (par. 2), een inhoudelijke richting (par. 3) en een richting die ziet op de werkwijze van annotatoren (par. 4). Een andere kwestie dan hoe annotaties te beoordelen, betreft de status van annotaties, meer in het bijzonder de vraag of zij als academisch product dienen te worden aangemerkt. Deze vraag vormt het slotstuk van deze studie (par. 5).

2 Procedurele waarborgen

De studies die zijn uitgevoerd, leveren een soms wat tegenstrijdig beeld op ten aanzien van de beoordeling van de kwaliteit van annotaties. De mate van overeenstemming tussen twee (of drie) beoordelaars varieerde tussen 50% en 86%. In de experimentele studies was de variatie over het algemeen gering. Het is niet duidelijk waar de verschillen precies aan moeten worden toegeschreven (vierpuntschaal versus zevenpuntschaal?, kritische beoordelaars versus positieve respondenten?), maar samengenomen suggereren de resultaten dat een beoordeling door meerdere beoordelaars kan leiden tot een redelijk nauwkeurige beoordeling van de kwaliteit van annotaties, zeker waar het gaat om een homogene groep van overwegend positieve beoordelaars.

Een beoordeling door meerdere beoordelaars verkleint de kans dat een annotatie niet wordt gepubliceerd terwijl deze positief wordt beoordeeld door ten minste een deel van het juridisch academisch forum, of andersom, dat een annotatie die door een groot deel van het juridisch forum als zwak of matig wordt aangemerkt, toch wordt gepubliceerd. Bovendien kan de feedback van beoordelaars dienen om de annotatie te verbeteren.

Sommige tijdschriften leggen geen redactionele toets aan. Zij selecteren gerenommeerde auteurs en veronderstellen dat deze auteurs behoorlijke annotaties leveren. Hoewel dat voor een deel ongetwijfeld het geval zal zijn, biedt deze vorm van kwaliteitsbewaking weinig tot geen garanties ten aanzien van de kwaliteit van een annotatie. Die garantie bestaat ook niet bij een beoordeling door meerdere beoordelaars en een redactionele toets, maar een dergelijke, serieuze beoordeling komt vermoedelijk wel ten goede aan de kwaliteit van annotaties.

Een beoordeling met meerdere beoordelaars draagt vermoedelijk bij aan het kunnen onderscheiden van goede annotaties van minder goede annotaties, maar een garantie is dit uiteraard nooit. Individuele missers zijn niet uit te sluiten, maar een systeem met nauwelijks of zelfs geen beoordeling vergroot de kans op missers. Bovendien dient de beoordeling gekoppeld te zijn aan inhoudelijke criteria.

3 Inhoudelijk criterium: normatieve diepgang

Voor de inhoudelijke beoordeling van annotaties keer ik terug naar Hoofdstuk 3, paragraaf 5, waarin het criterium van normatieve diepgang als overkoepelend criterium werd geïntroduceerd en is vormgegeven aan de hand van de gevonden kwaliteitsindicatoren. Die invulling gaf de contouren van het overkoepelende criterium, die hierna worden beschreven. Bij wat hierna volgt, past echter wel een waarschuwing. De criteria zijn louter een richtsnoer. Het is van belang open te blijven staan voor de mogelijkheid van nieuwe kenmerken en indicatoren. Er zullen altijd annotaties zijn waarin niet is voldaan aan een bepaald criterium en die toch als kwalitatief hoogstaand kunnen worden aangemerkt. Dergelijke uitzonderingen zijn echter moeilijk te duiden en zullen daarom nader dienen te worden gemotiveerd en onderbouwd.

Criterium 1 – De annotatie heeft een van onderstaande kenmerken

- *Implicaties van de uitspraak.* Het analyseren van de gevolgen gebeurt in de regel door de gewezen rechtsregel toe te passen op nieuwe, al dan niet zelf bedachte gevallen, aan de hand waarvan wordt beoordeeld of de regel wenselijk is, eventueel tegen de achtergrond van mogelijke alternatieven. Het schetsen van implicaties van een uitspraak kan aanleiding geven tot nuanceren of aanpassing van de gewezen regel. Een andere mogelijkheid is dat in de annotaties wordt bekeken of de uitspraak relevant is voor andere leerstukken, rechtsfiguren of rechtsgebieden.
- *Toetsen aan ‘hogere’ recht,* dat wil zeggen het toetsen van de gewezen uitspraak aan verdragsbepalingen, rechtsbeginselen of ander ‘hogere’ recht. Een mogelijkheid is om na te gaan of de uitspraak op gespannen voet staat met dit ‘hogere’ recht en, eventueel, om te bezien of een andere oplossing voor de hand had gelegen. Welk beginsel of welke verdragsbepaling een annotator kiest om de uitspraak aan te toetsen en, belangrijker, of die keuze voor de hand ligt, hangt mede af van het rechtsgebied of het leerstuk, alsook van de mate waarin een bepaald beginsel of verdragsbepaling als dominant wordt beschouwd binnen het betreffende rechtsgebied of leerstuk. Zo zal een beroep op de rechtszekerheid een grotere rol spelen binnen verjaring dan bij de uitleg van overeenkomsten. Indien meerdere beginselen of verdragsbepalingen

gen van toepassing zijn, mag worden verwacht dat de annotatie aangeeft in welke mate deze beginselen een rol spelen in de analyse.

- *Ontwikkelen van een criterium*, inclusief het vergelijken van dit criterium met de besproken uitspraak en/of het toepassen van dit criterium op één of meerdere situaties. Het criterium dat de annotator presenteert, zal dikwijls een concretisering vormen van een achterliggende waarde (zoals rechtszekerheid, legaliteitsbeginsel, etc.). Het ontwikkelen van een criterium draagt bij aan de ontwikkeling van het recht, meer concreet aan het vergroten van de voorspelbaarheid op het specifieke terrein waar de annotatie betrekking op heeft. Door het ontwikkelde criterium toe te passen op, of te vergelijken met, de uitspraak die in de annotatie centraal staat, kan de annotator toetsen of deze meer concrete criteria een adequate vertaling zijn van het achterliggende, algemene rechtsbeginsel (of rechtsbeginselen).
- *Het toetsen welke leer geldend recht is*. In dit soort gevallen gaat de annotator na of de uitspraak duidelijkheid schept over een bepaalde theorie of doctrine, bijvoorbeeld door te toetsen of een theorie die in de literatuur is ontwikkeld, wordt gevolgd in de uitspraak. Een andere mogelijkheid is dat er meerdere, concurrerende theorieën bestaan en dat wordt getoetst welke theorie een grotere verklarings- of voorspellingskracht heeft.
- De annotatie heeft kenmerken die vergelijkbaar zijn met een van de hiervoor genoemde kenmerken. Deze categorie is nodig, want de gevonden kwaliteitsindicatoren zijn niet limitatief. Bij het bestuderen van nieuwe of grotere aantallen annotaties kunnen zich nieuwe indicatoren aandienen. Te denken valt aan annotaties waarin, voor zover de rechter dat zelf nog niet heeft gedaan (!), gezichtspunten of criteria uit de uitspraak of uit een reeks van uitspraken worden afgeleid. Die gezichtspunten of criteria kunnen worden gebruikt om toekomstige gevallen te beslechten of om de uitkomst in die toekomstige gevallen te voorspellen.

Criterium 2 – De annotatie ontstijgt het individuele geval

De annotatie ontstijgt het individuele geval – de besproken uitspraak – wil zij als kwalitatief hoogstaand worden aangemerkt. Een sterke contra-indicatie hiervoor is als de annotatie in de kern uit niet veel anders bestaat dan:

- Het beschrijven of samenvatten van de feiten of de uitspraak.
- Het verschaffen van een overzicht van het geldende recht en literatuur daarover.
- Het schetsen van kwesties die zich geregeld voordoen.
- Het opmerken van motiveringsgebreken.
- Het ophelderen van onduidelijke passages uit de uitspraak, tenzij de implicaties van deze onduidelijkheden worden behandeld.

- Het gemotiveerd eens of oneens zijn met de uitspraak en/of de motivering van de uitspraak (anders dan door het schetsen van implicaties etc.). Het gaat bij dit kenmerk vooral om een kritische evaluatie van de uitspraak zelf, waarbij geen conclusies worden getrokken die bredere, normatieve implicaties hebben.
- Het opwerpen van (kritische) vragen, zonder deze adequaat te beantwoorden.
- Het toetsen of ter discussie stellen van opvattingen van andere auteurs, tenzij sprake is van het toetsen van bestaande (concurrerende) theorieën aan de uitspraak zoals hiervoor omschreven.

criterium 3 – Diepgang

In aanvulling op het voorgaande, dient de annotatie voldoende diepgang te hebben. Dat is niet het geval indien de uitspraak waarbij de annotatie is verschenen, wordt vergeleken met één of meerdere, eerder verschenen uitspra(a)k(en) door louter te constateren dat de besproken uitspraak overeenkomt of verschilt van die andere uitspra(a)k(en). Het is redelijk objectief vast te stellen of dit het geval is, ook voor degenen die minder goed zijn ingevoerd in het vakgebied.

Adequaat bronnengebruik is moeilijker vast te stellen voor wie minder goed thuis is in de materie. Het bronnengebruik moet zorgvuldig zijn, in ieder geval niet selectief. In het verlengde daarvan ligt dat indien de uitspraak wordt geïnterpreteerd of de gevolgen ervan worden ‘doorgerekend’, dit adequaat dient te gebeuren. Dit brengt mee dat de interpretatie realistisch moet zijn, en het doorrekenen reëel.

De omvang speelt in zoverre een rol, dat passages met kenmerken die positief gerelateerd zijn aan de kwaliteit van annotaties over het algemeen geen effect hebben op de kwaliteit als zij van zeer geringe omvang (enkele regels) zijn. Omgekeerd geldt dat een meer omvangrijke noot niet per definitie een goede noot is.

4 Criteria omtrent de werkwijze

Daar waar de gevonden kwaliteitsindicatoren in belangrijke mate inhoudelijk van aard zijn, zien de kenmerken van wetenschappelijkheid vooral op de manier waarop annotatoren te werk gaan (zouden moeten gaan?). Een manier om de eisen van diepgang en juistheid te operationaliseren of concretiseren is dan ook om annotaties (tevens) te beoordelen aan de hand van kenmerken van wetenschappelijkheid. Een nadeel van het gebruiken van kenmerken van wetenschap is echter dat zij niet specifiek zijn geschreven voor juridisch onderzoek en bovendien niet om individuele publicaties te beoordelen. Een vertaalslag is dan

ook nodig om kenmerken van wetenschappelijkheid te vertalen naar criteria van annotaties. Daartoe is hierna een poging gedaan:

- *Produceeren van systematische kennis.* De annotatie produceert kennis die generaliseerbaar is (ten minste het individuele geval ontstijgt). Dit komt in belangrijke mate overeen met de eerdere vaststelling dat goede/uitstekende annotaties bestaan uit analyses die het individuele geval ontstijgen, wat bij zwakke annotaties niet of minder vaak het geval is.
- *Zuiver bronnengebruik.* De vraag waar het hier om gaat, is of (vooral) bronnen en/of voorbeelden zijn gebruikt die het eigen standpunt of de opvatting van de annotator onderbouwen, of dat de annotator tevens voldoende oog heeft gehad voor andere opvattingen, standpunten en argumenten. Indien selectief gebruik is gemaakt van informatie, is dit een contra-indicatie voor een goede of uitstekende noot.
- *Repliceerbaarheid/verdedigbaarheid.* In een annotatie waarin de annotator een standpunt geeft, dient dit standpunt, alsook de onderliggende argumentatie, in redelijkheid verdedigbaar te zijn. Dit hoeft niet te betekenen dat de beoordelaar het eens is met de annotator, maar wel dat wat de annotator schrijft, verdedigbaar is, net zoals een natuurwetenschapper of sociale wetenschapper een door een collega gekozen methode verdedigbaar kan vinden terwijl hij of zij zelf een andere proefopstelling of manier van dataverzameling zou hebben gekozen. Het idee is dat een reviewer het niet eens hoeft te zijn met de gekozen invalshoek (bijvoorbeeld rechtszekerheid), maar wel moet kunnen nagaan hoe de annotator tot een bepaalde conclusie is gekomen en dus, gegeven de (geëxpliciteerde) invalshoek, zelf tot een vergelijkbare conclusie zou komen als de annotator. Met een criterium van 'verdedigbaarheid' kan tegemoet worden gekomen aan de diversiteit van de aanpak binnen annotaties. Wel vereist dit criterium dat annotatoren het perspectief, de invalshoek of het beginsel van waaruit zij redeneren, voldoende concreet maken en bovendien expliciteren. Zo zal het legaliteitsbeginsel in het strafrecht of het publicatiebeginsel in het goederenrecht eerder voldoen aan het vereiste van concreetheid dan een beginsel als de redelijkheid en billijkheid. Overigens zal geen sprake kunnen zijn van een binair of dichotoom onderscheid tussen repliceerbaar en niet-repliceerbaar. Een glijdende schaal ligt meer voor de hand.
- *Kritiek met alternatief.* Indien de annotator kritiek heeft op de uitspraak van de rechter, mag in een kwalitatief hoogstaande annotatie van de annotator worden verwacht dat deze een alternatief aandraagt dat de kritiek ondervangt en/of een betere oplossing biedt dan de uitspraak. Hieronder kan tevens worden begrepen het zich kritisch uitlaten over de uitspraak, gevolgd door de vaststelling dat alternatieven geen betere oplossing bieden dan de uitspraak in kwestie.

De benadering waarbij de indicatoren voor wetenschappelijkheid (mede)bepalend zijn voor de kwaliteit van annotaties kan belangrijke consequenties hebben. De belangrijkste is dat annotaties die thans als goed of uitstekend worden aangemerkt, niet per definitie diezelfde kwalificatie krijgen in een beoordelingssysteem waarbij de hiervoor criteria worden betrokken. Aan de andere kant suggereren de resultaten van het uitgevoerde experiment dat als beoordelaars wetenschappelijke criteria krijgen aangereikt, dit hun beoordelingen niet wezenlijk beïnvloedt. Dit inzicht komt echter voort uit een studie met weinig respondenten die bovendien slechts één annotatie hebben gelezen en hebben beoordeeld.

Het is reëel te verwachten dat een deel van de annotaties er in de toekomst anders uit komt te zien als de hiervoor genoemde criteria van wetenschappelijkheid (mede) worden gebruikt om annotaties te beoordelen. In annotaties hebben annotatoren geregeld kritiek op de uitspraak of de motivering ervan. Minder gebruikelijk is het om alternatieven voor de gewezen beslissing of de motivering ervan te bespreken. Op grond van voorgaande criteria is een andere aanpak zuiverder. Die voert naar het bespreken van alternatieven en het schetsen van de voor- en nadelen van die alternatieven in relatie tot de genomen beslissing. Een rechterlijke uitspraak waarin een verjaringstermijn in een asbestzaak is doorbroken, zou op deze manier, heel in het kort, ongeveer als volgt worden besproken:¹⁵³

De rechter heeft in casu de verjaringsregel niet toegepast. Een alternatief zou zijn geweest om vast te houden aan de verjaringstermijn en deze toe te passen. Daar pleit voor dat de ratio van verjaring is dat men na verloop van tijd niet meer geconfronteerd hoeft te worden met claims. Dit in verband met mogelijke bewijsproblemen waar partijen zich mee geconfronteerd zien, maar ook in verband met het meer principiële punt dat men na verloop van tijd geen claims meer hoeft te verwachten. Aan de andere kant doet zich in casu een bijzondere omstandigheid voor, namelijk dat de gelaedeerde pas laat (na het verstrijken van de verjaringstermijn) bekend is geworden met de schade en dus geen reële mogelijkheid heeft gehad om de schade te vorderen. Dit pleit vóór de uitspraak. Een gevaar van deze redenering zou kunnen zijn dat de verjaringstermijn op grotere schaal opzij wordt geschoven. Of dit risico reëel is, valt te betwijfelen, omdat het aantal gevallen waarin de schade pas intreedt nadat een verjaringstermijn is verstreken, beperkt lijkt. [etc.]

Wat bovenstaande, summiere, fictieve annotatie anders maakt dan veel van de bestaande annotaties, is dat niet zozeer standpunten worden ingenomen, maar verschillende mogelijke oplossingen worden besproken, worden doorgerekend en worden vergeleken. Een dergelijke aanpak komt ten goede aan de repliceer-

¹⁵³ Bij wijze van voorbeeld. Bronvermelding en uitputtendheid van argumenten ontbreken.

baarheid: voor anderen dan de annotator is het mogelijk om na te lopen of de annotator de juiste argumenten heeft gebruikt, of deze zijn gebaseerd op de juiste bronnen en of hij/zij alternatieven over het hoofd heeft gezien. Het gaat bij het schrijven en beoordelen van een goede annotatie dan niet zozeer om het innemen van een standpunt in termen van een voorkeur voor een bepaalde oplossing, maar om, veel neutraler, inzicht te geven in de redenen voor en consequenties van de beslissing.

5 Status van annotaties

Van de vraag of een annotatie als academisch product *kan* worden aangemerkt dient te worden onderscheiden de vraag of een annotatie als academisch product *moet* worden aangemerkt. Zo kunnen annotaties als niet- of minder wetenschappelijk worden aangemerkt, maar toch als nuttig en zinvol worden ervaren.

Voor het antwoord op de laatste vraag spelen ook andere elementen mee dan alleen de kenmerken van annotaties of kenmerken van wetenschappelijkheid. Ook van belang zijn de institutionele inbedding van annotaties (motiveren rechters hun beslissingen al uitvoerig?, hebben annotaties impact op de ontwikkeling van het recht?, wat voegen annotaties daar aan toe?) en de verhouding van de annotatie tot andere publicatievormen, zoals juridische artikelen en dissertaties. Relevant is tevens de richting waarin het juridische onderzoek (in Nederland) zich beweegt. Ook dit bepaalt mede de status die annotaties zouden dienen te krijgen. In deze paragraaf tracht ik dit bredere kader te duiden, ga ik na hoe (goede) annotaties hierin passen en formuleer ik (voorzichtig) een antwoord op de vraag of de annotatie de status van academisch product zou moeten hebben.

In Hoofdstuk 1, paragraaf 1 verwees ik naar verschillende auteurs die de waarde van annotaties hebben bepleit. Zij hebben hun waardering uitgesproken over de aard van de annotaties en de kennisvergroting die zij meebrengen.¹⁵⁴ In dezelfde paragraaf besprak ik ook waarom de annotatie als publicatievorm onder druk staat. Deze argumenten, die ik hierna uitwerk en aanvul, geven aan dat een belangrijk deel van de juridische onderzoekers twijfels heeft over het nut en de relevantie van annotaties.

Allereerst is er het rapport Oordelen over rechten (2005) van de *Commissie Voorbereiding Onderzoeksbeoordeling Rechtsgeleerdheid*. Dat noemt in de kern een drietal redenen om annotaties niet als academische publicatie te beschouwen. De eerste is dat annotaties divers van aard zijn en daardoor niet altijd de vereiste originaliteit hebben. Annotaties die wel origineel zijn, zullen vaak ook worden uitgewerkt in een andere publicatie, zoals in een artikel. De tweede

¹⁵⁴ Bijvoorbeeld Snijders 2003; Jansen 2003; Bloembergen 1988; Bloembergen 2002.

reden is dat annotaties qua opzet, omvang en onderzoek- en schrijftijd afwijken van bijvoorbeeld dissertaties, boeken en artikelen. De derde reden die het rapport geeft, is dat een goede maatstaf ontbreekt om annotaties te beoordelen. Veel juridische tijdschriften hebben een annotatierubriek, waarvan de aard en kwaliteit sterk kan wisselen.

De overtuigingskracht van de eerste en laatste reden uit het rapport, die beide in essentie neerkomen op het ontbreken van beoordelingscriteria, zou na deze studie moeten zijn afgenomen. Daarbij komt dat de onderbouwing niet sterk is, want mogelijk geldt veel van wat wordt aangevoerd ook voor artikelen in juridische tijdschriften. Bovendien kan de geringe omvang van een annotatie, in vergelijking met bijvoorbeeld een artikel, relatief eenvoudig worden gecompenseerd door er een gewicht ('impactfactor') aan toe te kennen dat recht doet aan de tijdsinvestering in relatie tot andere publicatievormen (artikelen, dissertaties, boeken en bundelbijdragen). Hoe dit ook zij, uit het rapport blijkt duidelijk de mindere waardering voor annotaties in vergelijking met andere publicatievormen.

Belangrijk zijn voorts de verschillen die er bestaan tussen annotaties en juridisch dissertatieonderzoek. In annotaties staat vaak het rolmodel van de rechter centraal. Dit brengt mee dat de annotator zich van dezelfde (interpretatie)methoden bedient als de rechter, wat zich vertaalt in een gebondenheid aan het geldende recht.¹⁵⁵ Deze gebondenheid heeft als gevolg dat oplossingen voor gesignaleerde problemen (dienen te) worden opgelost binnen het stelsel van de wet, wetsgeschiedenis en jurisprudentie.

Onderzoek laat zien dat het rolmodel van de rechter dat in annotaties centraal staat, in beperkte mate wordt gevolgd in juridisch dissertatieonderzoek.¹⁵⁶ In dissertaties is niet het rolmodel van de rechter dominant, maar dat van de wetgever. Binnen dit rolmodel evalueert de juridisch onderzoeker een regel of zoekt hij naar de meest wenselijke oplossing.¹⁵⁷ In tegenstelling tot bij het rolmodel van de rechter, is de juridisch onderzoeker binnen het rolmodel van de wetgever in mindere mate gebonden aan het geldende recht. In dit verband kan worden gesproken van een 'extern perspectief'.

Het contrast tussen juridisch dissertatieonderzoek en annotaties kan een aanwijzing zijn voor een verschil in waardering. Wie dissertaties beschouwt als belangrijke en volwaardige academische producten die richtinggevend zijn voor het juridisch onderzoek dat wordt verricht, kan concluderen dat in annotaties iets

¹⁵⁵ Zie hierover De Geest 2004, p. 62-63; Vranken 2005, nr. 144; Westerman & Wissink 2008; Van Dijk & Van Gulijk & Prinsen 2010.

¹⁵⁶ Van Dijk & Van Gulijk & Prinsen 2010.

¹⁵⁷ Van Dijk & Van Gulijk & Prinsen 2010, p. 51.

wezenlijk anders gebeurt dan in dissertatieonderzoek. Deze vaststelling kan impliciet een oordeel inhouden over de (lagere) kwaliteit van annotaties.¹⁵⁸

Daarnaast is van belang te kijken naar de verhouding tussen uitspraken en annotaties. In vergelijking met de hoogste Franse rechter, motiveren de Hoge Raad en de Raad van State in Nederland hun uitspraken uitvoerig. Ook in de conclusie van de A-G (voorheen P-G) zijn beschouwingen te vinden over wat de uitspraak zou moeten zijn en waarom, en vermelden bovendien relevante literatuur, jurisprudentie en argumenten. Dit heeft tot gevolg dat rechters en het parket voor een deel het gras voor de voeten van de annotator wegmaaien, zeker in vergelijking met een rechtsstelsel als dat van Frankrijk, en dat het daarom relatief moeilijk is voor annotatoren om een punt te maken. Dit beïnvloedt de waardering van annotaties. Logischerwijs zal een annotatie die herhaalt wat de rechter (of een A-G) heeft gesteld minder waardering oogsten dan een annotatie die ‘iets’ toevoegt ten opzichte van de uitspraak.

Het is overigens moeilijk vast te stellen of, en zo ja, waarom annotaties vroeger meer werden gewaardeerd dan thans. Niet alleen zagen gerechtelijke uitspraken er toen anders uit (minder gemotiveerd), maar ook annotaties hadden een andere vorm en inhoud. Het ligt voor de hand dat de plaats en inhoud van annotaties in samenhang moet worden gezien met de ontwikkeling in de motivering van uitspraken. Daarnaast spelen ook subsidiegevers een rol. Deze, met name NWO, lijken of zeggen vooral belangstelling te hebben voor onderwerpen die het nationale overstijgen zijn en/of een multidisciplinaire inslag hebben.¹⁵⁹ Rechtenfaculteiten nemen deze eisen over bij het toekennen van onderzoeksgelden. Hierdoor lijken annotaties het onderspit te delven. Deze zijn immers overwegend nationaal georiënteerd, in ieder geval in rechtsgebieden als het bestuursrecht, privaatrecht en strafrecht.

Duidelijk in ieder geval is dat er binnen het juridisch-academisch forum geen overeenstemming bestaat over het nut en de relevantie van annotaties. Of de dagen van annotaties als academisch product zijn geteld, hangt echter uiteindelijk af van de waardering van de annotatie door het juridisch-academisch forum. Feitelijke informatie die deze studie biedt is daarbij belangrijk en noodzakelijk, maar kan op zichzelf niet doorslaggevend zijn voor het beantwoorden van de normatieve vraag over de status van annotaties.

¹⁵⁸ Het is moeilijk na te gaan of de verschillen tussen annotaties en juridisch dissertatieonderzoek vroeger anders waren. Dat zou empirisch onderzocht kunnen worden door dissertatieonderzoek uit bijvoorbeeld 1960 te vergelijken met annotaties uit die periode en eventueel met hedendaags dissertatieonderzoek.

¹⁵⁹ Toen ik bij NWO naging welke projecten in de open competitie aan juristen zijn gehonoreerd, viel interessant genoeg op dat met name internationale onderwerpen (bijvoorbeeld mensenrechten) met een ogenschijnlijk klassiek-juridische invalshoek geregeld in de prijzen zijn gevallen.

Overzicht van verkort aangehaalde literatuur

Barendrecht 1996

J.M. Barendrecht, 'Rechtswetenschap: stoffig of inventief? De Nederlandse universitaire rechtswetenschap beoordeeld door een visitatiecommissie', *NJB* 1996, p. 705-714.

Bell, Boyron & Whittaker 1998

John Bell, Sophie Boyron & Simon Whittaker, *Principles of French Law*, Oxford: Oxford University Press 1998.

Bleeker 2010

R.G.T. Bleeker, 'De noot', in: M.A.B. Chao-Duvis, C.E.C. Jansen & J.B.M. Vranken, *Alleen samen: Opstellen aangeboden aan prof. mr. M.A.M.C. van den Berg*, Den Haag: Instituut voor Bouwrecht 2010, p. 483-492.

Bloembergen 1988

A.R. Bloembergen, 'Iets over object en methode van wetenschap en rechtspraak in het privaatrecht', in: O.W.M. Kamstra & F.B.M. Kunneman, *Nederlandse rechtswetenschap: tussen distantie en betrokkenheid: paradigma's in de twintigste eeuw*, Zwolle: Tjeenk Willink 1988, p. 61-80.

Bloembergen 2002

A.R. Bloembergen, 'Over het annoteren van arresten', *NTBR* 2002, p. 88-99.

Bunge 1982

Mario Bunge, 'Demarcating Science from Pseudoscience', *Fundamenta Scientiae* 1982, p. 369-388.

Calleros 2001

Charles R. Calleros, 'Using Classroom Demonstrations in Familiar Nonlegal Contexts to Introduce New Students to Unfamiliar Concepts of Legal Method and Analysis', *Legal Writing: J. Legal Writing Inst.* 2001, p. 37-62.

Crombag & Cohen 1974

H.F. Crombag & M.J. Cohen, *Over het annoteren van arresten: Bloembergen over het fluorideringsarrest*, Leiden: Rijksuniversiteit Leiden, bureau onderzoek van onderwijs 1974.

De Geest 2004

G. de Geest, 'Hoe maken we van de rechtswetenschap een volwaardige wetenschap?', *NJB* 2004, p. 58-66.

Degenkamp 2004

J.Th. Degenkamp, 'Reacties - Franken: onverbeterlijk?', *NJB* 2004, p. 2246.

Derksen 1993

A.A. Derksen, 'The seven sins of pseudo-science', *Journal for General Philosophy of Science* 1993, p. 17-42.

Derksen 2003

Anthony A. Derksen, 'The Seven Strategies of the Sophisticated Pseudo-Scientist: a look into Freud's rhetorical tool box ', *Journal for General Philosophy of Science* 2003, p. 329-350.

Dutch 1982

Steven I. Dutch, 'Notes on the nature of fringe science', *Journal of Geological Education* 1982, p. 6-13.

Duxbury 2001

Neil Duxbury, *Jurists and Judges: An Essay on Influence*, Oxford: Hart Publishing 2001.

Feyerabend 1975

Paul Feyerabend, *Against Method*, London: NLB 1975.

Franken 2004

H. Franken, 'Rechtsgeleerdheid in de rij der wetenschappen', *NJB* 2004, p. 1400-1408.

Gionfriddo 2007

Jane Kent Gionfriddo, 'Thinking Like a Lawyer: The Heuristics of Case Synthesis', *Texas Tech Law Review* 2007, p. 1-36.

Glymour & Stalker 1990

Clark Glymour & Douglas Stalker, 'Winning through Pseudoscience', in: Patrick Grim, *Philosophy of Science and the Occult*, Albany: State University of New York Press 1990, p. 140-155.

Goutal 1976

Jean Louis Goutal, 'Characteristics of Judicial Style in France, Britain and the U.S.A.', *Am. J. Compar. Law* 1976, p. 43-72.

Greenstein 1985-1986

Richard K. Greenstein, 'Teaching Case Synthesis', *GA. ST. U. L. REV.* 1985-1986, p. 1-17.

Grove 1985

J.W. Grove, 'Rationality at Risk: Science against Pseudoscience', *Minerva* 1985, p. 216-240.

Gruenberger 1964

Fred J. Gruenberger, 'A measure for crackpots', *Science* 1964, p. 1413-1415.

Gutting 2000

Gary Gutting, 'Scientific Methodology', in: W.H. Newton-Smith, *A Companion to the Philosophy of Science*, Malden: Blackwell 2000, p. 423-432.

Hansson 1983

Sven Ove Hansson, *Vetenskap och ovetenskap*, Stockholm: Tiden 1983.

Hansson 2003

Sven Ove Hansson, 'Science and Pseudo-Science', <<http://plato.stanford.edu/entries/pseudo-science/>>.

Hesselink 2009

Martijn Hesselink, 'A European Legal Method? On European Private Law and Scientific Method', *European Law Journal* 2009, p. 20-45.

Hondius 2010

Ewoud Hondius, 'Over het annoteren van arresten / naar aanleiding van HR 18 juni 2010, LJN BM0893 m.nt. E.S. Daalder', in: A.G. Castermans e.a., *40 jaar practicum. Randon onroerend goed*, Deventer: Kluwer 2010, p. 144-148.

Jansen 2003

C.J.H. Jansen, 'De Nederlandse bakermat van het annoteren', *NJB* 2003, p. 1557-1559.

Kitcher 1982

Philip Kitcher, *Abusing Science. The Case Against Creationism*, Cambridge, MA: MIT Press 1982.

Kranenburg 1955

R. Kranenburg, *De grondslagen der rechtswetenschap: juridische kennisleer en methodologie*, Haarlem: Tjeenk Willink 1955.

Kuhn 1962

Thomas S. Kuhn, *The Structure of Scientific Revolutions*, Chicago & London: The University of Chicago Press 1962.

Kuhn 1977

Thomas S. Kuhn, *The essential tension: selected studies in scientific tradition and change*, Chicago: The University of Chicago Press 1977.

Lakatos 1978

Imre Lakatos, *The methodology of scientific research programmes*, Cambridge: Cambridge University Press 1978.

Landis & Koch 1977

J.R. Landis & G.G. Koch, 'The measurement of observer agreement for categorical data', *Biometrics* 1977, p. 159-174.

Langmuir 1989 (1953)

Irving Langmuir, 'Pathological Science', *Physics Today* 1989 (1953), p. 36-48.

Latour 1987

Bruno Latour, *Science in Action*, Cambridge, MA: Harvard University Press 1987.

Laudan 1983

Larry Laudan, 'The Demise of the Demarcation Problem', in: R.S. Cohen & L. Laudan, *Physics, Philosophy and Psychoanalysis: Essays in Honor of Adolf Grünbaum*, Dordrecht: D. Reidel 1983, p. 111-127.

Laudan 1984

Larry Laudan, *Science and Values: the Aims of Science and their Role in Scientific Debate*, Berkeley / Los Angeles: Berkeley University Press 1984.

Lundstedt 1930/1936

A.V. Lundstedt, *Die Unwissenschaftlichkeit der Rechtswissenschaft*: 1930/1936.

Mahner 2007

Martin Mahner, 'Demarcating Science from Non-Science', in: Theo Kuipers, *Handbook of the Philosophy of Science: General Philosophy of Science – Focal Issues*, Amsterdam: Elsevier 2007, p. 515-575.

McGrew, Alspector-Kelly & Allhoff 2009

Timothy McGrew, Marc Alspector-Kelly & Fritz Allhoff, *Philosophy of Science: An Historical Anthology*, Malden, MA: Wiley-Blackwell 2009.

Montana-Law-Review 2008

Montana-Law-Review, 'Suggested Structure of a Law Review Note', <<http://www.montanalawreview.com/sitebuildercontent/sitebuilderfiles/casenoteoutline.doc>>.

Oosten 2004

Kees van Oosten, 'Recht als wetenschap (Reactie op G. de Geest, 'Hoe maken we van de rechtswetenschap een volwaardige wetenschap?', *NJB* 2004/2, p. 58-66)', *NJB* 2004, p. 1436.

Peterson 2003

Gregory R. Peterson, 'Demarcation and the Scientistic Fallacy', *Zygon: Journal of Religion and Science* 2003, p. 751-761.

Popper 1974

Karl R. Popper, *Conjectures and refutations: the growth of scientific knowledge*, London: Routledge & Kegan Paul 1974.

Popper 1980

Karl R. Popper, *The logic of scientific discovery*, London: Hutchinson 1980.

Popper 1983

Karl R. Popper, *Realism and the aim of science*, Totowa, NJ: Rowman and Littlefield 1983.

Radner & Radner 1982

Daisie Radner & Michael Radner, *Science and Unreason*, Belmont, CA: Wadsworth 1982.

Ragin 2000

Charles C. Ragin, *Fuzzy-Set Social Science*, Chicago: University of Chicago Press 2000.

Ragin 2006

Charles C. Ragin, 'Set Relations in Social Research: Evaluating Their Consistency and Coverage', *Political Analysis* 2006, p. 291-310.

Rihoux & Ragin 2008

Benoît Rihoux & Charles C. Ragin, *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, Thousand Oaks: SAGE Publications 2008.

Ruse 1996

Michael Ruse, *But is it science? The philosophical question in the creation/evolution controversy*: Prometheus Books 1996.

Singleton & Straits 2005

Royce A. Singleton & Bruce C. Straits, *Approaches to Social Research*, New York/Oxford: Oxford University Press 2005.

Smith e.a. 2008

C.E. Smith e.a., 'Criteria voor goed rechtswetenschappelijk onderzoek. De omgekeerde route', *NJB* 2008, p. 685-690.

Smith 1998-1999

Kevin H. Smith, 'Practical Jurisprudence: Deconstructing and Synthesizing the Art and Science of Thinking like a Lawyer', *U.MEM. L. REV.* 1998-1999, p. 1-68.

Smits 2009

Jan Smits, *Omstreden rechtswetenschap: Over aard, methode en organisatie van de juridische discipline*, Den Haag: Boom Juridische uitgevers 2009.

Snijders 2003

H.J. Snijders, 'Annoteren, een summum van rechtswetenschappelijk presteren', *NTBR* 2003, p. 173.

Snijders & Rank-Berenschot 2007

H.J. Snijders & E.B. Rank-Berenschot, *Goederenrecht*, Deventer: Kluwer 2007.

Stolker 2003

C.J.J.M. Stolker, 'Ja, geléerd zijn jullie wel! Over de status van de rechtswetenschap', *NJB* 2003, p. 766-778.

Stolker 2004

C.J.J.M. Stolker, 'Wat maakt een juridisch tijdschrift wetenschappelijk?', *NJB* 2004, p. 1409-1418.

Thagard 1988

Paul R. Thagard, *Computational Philosophy of Science*, Cambridge, MA: MIT Press 1988.

Tijssen 2009

Hervé Tijssen, *De juridische dissertatie onder de loep: De verantwoording van methodologische keuzes in juridische dissertaties (diss. Tilburg)*, Tilburg: Boom Juridische Uitgevers 2009.

Tjittes 2005

Rieme-Jan Tjittes, 'Hoe verkeerd doen de civielrechtelijke wetenschap en rechtspraak het eigenlijk?', *NJB* 2005, p. 1879-1884.

Tunc 1976

André Tunc, 'Methodology of the Civil Law in France', *Tulane L. Rev.* 1976, p. 459 e.v.

Van Dijk, Van Gulijk & Prinsen 2010

Gijs Van Dijk, Stéphanie Van Gulijk & Merel Prinsen, 'Wat doen juridische onderzoekers? Een empirische blik', *Recht der Werkelijkheid* 2010, p. 44-64.

Van Gestel & Vranken 2007

R.A.J. Van Gestel & J.B.M. Vranken, 'Rechtswetenschappelijke artikelen. Naar criteria voor methodologische verantwoording', *NJB* 2007, p. 1448-1461.

Van Rhee 2004

C.H. Van Rhee, 'Geen rechtsgeleerdheid, maar rechtswetenschap!', *NTBR* 2004, p. 196-201.

Vollmer 1993

Gerhard Vollmer, *Wissenschaftstheorie im Einsatz, Beiträge zu einer selbstkritischen Wissenschaftsphilosophie*, Stuttgart: Hirzel Verlag 1993.

Vranken 2005

J.B.M. Vranken, *Algemeen deel***: een vervolg*, Deventer: Kluwer 2005.

Wendt 2008

J.A.I. Wendt, *De methode der rechtswetenschap vanuit kritisch-rationeel perspectief* (diss. Rotterdam), Zutphen: Paris 2008.

Wendt 2009

J.A.I. Wendt, 'De opgekomen methodenvrees in het rechtswetenschappelijk debat in Nederland. Een voorstel', *NJB* 2009, p. 782-789.

Westerman & Wissink 2008

P. Westerman & M. Wissink, 'Rechtsgeleerdheid als rechtswetenschap', *NJB* 2008, p. 503-507.

Betrouwbaarheid vragenlijst & selectie annotaties

1 Meten van betrouwbaarheid

In het algemeen geldt dat een meetinstrument betrouwbaar is indien verschillende onderzoekers het op dezelfde wijze toepassen. Voor een hoge betrouwbaarheid is ten minste een goede operationalisering (of concretisering) van de begrippen vereist. Indien deze achterwege blijft, bestaat immers het risico dat de begrippen op verschillende manieren zullen worden geïnterpreteerd of opgevat. Daarmee is het meetinstrument onbetrouwbaar.

De ingevulde vragenlijsten voor de privaatrechtelijke annotaties zijn gebruikt om de betrouwbaarheid van de vragenlijst te meten. De vragenlijst spitste zich toe op de kenmerken van annotaties, het bronnengebruik en de (praktische en theoretische) relevantie. Door met deze vragenlijst proef te draaien, is met een aantal collega-onderzoekers verkend of de begrippen voldoende helder zijn omschreven en, daarmee, of de vragenlijst voldoende duidelijke instructies geeft. Waar dit niet het geval was, is de vragenlijst aangescherpt totdat er geen interpretatieverschillen meer werden waargenomen.

2 Selectie van annotaties en beoordelaars

De vragenlijst bevat vragen met vier- en vijfpuntschalen. Om de betrouwbaarheid te meten, zijn de schalen teruggebracht twee antwoordcategorieën (bijvoorbeeld (helemaal) niet versus (helemaal) wel, of nee/(heel) weinig versus (heel) veel). In totaal zijn 16 (privaatrechtelijke) annotaties geselecteerd. Met betrekking tot het tellen van het aantal woorden is een betrouwbaarheidstoets niet nodig en daarom achterwege gelaten.

De selectie van annotaties heeft plaatsgevonden door aan een deskundige op het gebied van het privaatrecht te vragen een hoeveelheid annotaties te selecteren, waarvan de helft van goede/uitstekende kwaliteit en de andere helft van matige/slechte kwaliteit. Tevens is aan deze expert gevraagd om het verschil in kwaliteit te verklaren. In dergelijke gevallen kan worden nagegaan of de expert die de annotaties aanreikt, een andere selectie heeft toegepast dan degenen die de annotaties analyseren, over wie hierna meer. De motivering geeft mogelijk, zo was de gedachte, verklaringen voor situaties waarin het verschil in kwaliteit niet tot uitdrukking komt in de antwoorden op de vragen uit de vragenlijst.

Voor het vaststellen van de betrouwbaarheid van de vragen zijn vervolgens twee andere deskundigen op het vakgebied benaderd. Deze deskundigen

zijn beiden werkzaam binnen de academische wereld (1x UHD, 1x hoogleraar). Het selecteren van vooraanstaande onderzoekers binnen de verschillende vakgebieden sluit aan bij het doel van deze studie, namelijk het ontwikkelen van criteria om de *academische* kwaliteit van annotaties te beoordelen. Zij kunnen het bronnengebruik en de relevantie het beste beoordelen. De deskundigen waren geen verklaarde voor- of tegenstanders van annotaties, dat wil zeggen deskundigen die op voorhand heel positief of juist heel negatief tegenover annotaties staan.

3 Uitkomst betrouwbaarheidsonderzoek

Indien een beoordelaar het antwoord ‘niet van toepassing’ heeft aangevinkt, is dit antwoord (en dus ook dat van de andere beoordelaar) buiten beschouwing gelaten bij het uitvoeren van de hiernavolgende analyse. Deze situatie doet zich met name voor bij vervolgvragen. Als bijvoorbeeld het antwoord ‘nee’ is gegeven op de vraag of de annotator een standpunt inneemt, kruist de beoordelaar ‘n.v.t.’ of een vergelijkbaar antwoord aan bij de vraag of de annotator het standpunt onderbouwt. Een enkele keer komt het voor dat een bepaalde vraag steeds met ‘n.v.t.’ of een vergelijkbaar antwoord is beantwoord. Die antwoorden komen niet terug in Tabel 15.

Tabel 15 – Intercoder reliability tests

Categorie	Overeenstemming tussen beoordelaars	Kappa-waardes
Leesbaarheid	53% (8/15)	.12
Opbouw/toegankelijkheid	60% (9/15)	.29
Bronnengebruik		
Wet (binnen rechtsgebied)	80% (12/15)	.33†
Wet (interne rechtsvergelijking)	100% (15/15)	---
Wet (externe rechtsvergelijking)	100% (15/15)	---
Wetsgeschiedenis (binnen rechtsgebied)	93% (14/15)	.63**
Wetsgeschiedenis (interne rechtsvergelijking)	100% (15/15)	---
Wetsgeschiedenis (externe rechtsvergelijking)	100% (15/15)	---
Jurisprudentie (binnen rechtsgebied)	80% (12/15)	.61**
Jurisprudentie (interne rechtsvergelijking)	100% (15/15)	1.00***
Jurisprudentie (externe rechtsvergelijking)	100% (15/15)	---
Literatuur (binnen rechtsgebied)	73% (11/15)	.25
Literatuur (interne rechtsvergelijking)	100% (15/15)	---
Literatuur (externe rechtsvergelijking)	93% (14/15)	---
Literatuur (overig)	100% (15/15)	---
Andere bronnen	100% (15/15)	---
Gebrek aan bronnen	45% (5/11)	.03
Correct bronnengebruik	87% (13/15)	---
Essentiële vragen/elementen	73% (11/15)	-.15
Nieuwe problematiek in uitspraak	79% (11/14)	.55*

Nieuw problematiek in annotatie (ook t.o.v. uitspraak)	100% (1/1)	---
Grootte van het probleem in uitspraak	67% (10/15)	.19
Grootte van het probleem nieuw aangestipt in annotatie (ook t.o.v. uitspraak)	100% (1/1)	---
Feitelijke informatie	100% (15/15)	---
Signaleren vraag- of knelpunten	69% (9/13)	---
Constructieve bijdrage aan oplossen probleem	0% (0/1)	---
Bespreking voorstellen/standpunten van anderen	93% (13/14)	---
Weerleggen standpunten anderen	33% (1/3)	---
Praktische relevantie uitspraak	53% (8/15)	.16
Praktische relevante annotatie	64% (9/14)	---

Noot: * $p < .05$. ** $p < .01$. † $p < .10$.

Gelet op de percentages, is sprake van een behoorlijke mate van overeenstemming voor veel categorieën. Dit geldt met name voor het bronnengebruik. Voor zover het berekenen van kappa-waardes mogelijk was en de analyses significante resultaten opleverden, variëren de waardes van zwak/matig (.30) tot middelgroot (.60) tot zeer groot (>.80). De kappa-waarde bleek nogal eens niet te kunnen worden berekend omdat de antwoorden van één of beide beoordelaars niet variëren. De beoordelaar geeft op een bepaalde vraag bijvoorbeeld steeds het antwoord 'niet/(heel) weinig', zodat het antwoord '(heel) veel' niet wordt gegeven en er geen kappa-waarde kan worden berekend. Soms blijken berekende kappa-waardes ook zeer laag en bovendien niet significant. Dit geldt bijvoorbeeld voor de vragen die betrekking hebben op het gebrek aan bronnen, essentiële vragen/elementen, de reikwijdte van het probleem in de uitspraak en de praktische relevantie. Een mogelijke verklaring hiervoor is de omstandigheid dat de vragen mede om een waardering vroegen. Er is dan ook meer overeenstemming over de meer objectieve vragen over welke bronnen er worden gebruikt dan over bijvoorbeeld de relevantie.

Overzicht scores potentiële kwaliteitsindicatoren

Drie deskundigen, waarvan twee privaatrechtelijke en één strafrechtelijke, hebben een vragenlijst ingevuld die bestond uit vragen die mogelijke kwaliteitsindicatoren vertegenwoordigden. De beoordelaars dienden hun scores op een vierpuntschaal (1-2-3-4) tot uitdrukking te brengen. Er is gewerkt met een vierpuntschaal omdat deze het mogelijk zou maken om de betrouwbaarheid van de vragenlijst te meten.¹⁶⁰

Een serie van T-toetsen is gebruikt om per beoordelaar na te gaan of er verschillen zouden bestaan tussen de annotaties die als goed/uitstekend zijn aangemerkt en die als matig/slecht zijn gekwalificeerd door de beoordelaars. Tabel 16 bevat de resultaten van de uitgevoerde toetsen.

Tabel 16 – Verschillen scores tussen matige/slechte annotaties en goede/uitstekende annotaties

Criterium	Beoordelaar	Kwaliteit annotatie		F	p	d
		Matig/slecht	Goed/uitstekend			
Aantal woorden	A (Pr)	2160	2974.33	F(1,13) =1.887	.19	.25
	B (Pr)	2473	2520.25	F(1,13) =.005	.95	.05
	C (Sr)	1275	1950	F(1,14) =2.199	.16	.28
Leesbaarheid	A (Pr)	3.44	4.00	F(1,13) = 1.757	.21	.23
	B (Pr)	2.73	2.50	F(1,13) =.115	.74	.06
	C (Sr)	2.25	2.92	F(1,14) =1.191	.29	.08
Op- bouw/toegankelijk- heid	A (Pr)	2.89†	3.83†	F(1,13) =4.294	.06	.48
	B (Pr)	2.55	3.00	F(1,13) =.903	.36	.14
	C (Sr)	2.75†	3.17†	F(1,14) =3.107	.10	.37
Wet (binnen rechts- gebied)	A (Pr)	1.22	1.17	F(1,13) =.012	.92	.05
	B (Pr)	1.64	1.75	F(1,13) =.023	.88	.05

¹⁶⁰ De scores op de vierpuntschaal zijn in dat verband teruggebracht tot dichotome scores, die vervolgens zijn gerelateerd aan de kwaliteitsoordelen, die tevens op een vierpuntschaal zijn gemeten en zijn teruggebracht tot een dichotome variabele. Zie Bijlage 1 voor de resultaten.

	C (Sr)	.75	1.00	F(1,14) =.127	.73	.06
Wet (interne rechts- vergelijking)	A (Pr)	.33	.50	F(1,13) =.173	.68	.07
	B (Pr)	.36	.50	F(1,13) =.094	.76	.06
	C (Sr)	.25†	.00†	F(1,14) =3.500	.08	.41
Wet (externe rechtsvergelijking)	A (Pr)	.11	.17	F(1,13) =.084	.78	.06
	B (Pr)	.18	.00	F(1,13) =.347	.57	.09
	C (Sr)	.00	.25	F(1,14) =.618	.45	.11
Wetgeschiedenis (binnen rechtsge- bied)	A (Pr)	.78	.50	F(1,13) =.399	.54	.09
	B (Pr)	.73	1.25	F(1,13) =.454	.51	.10
	C (Sr)	.25	.08	F(1,14) =.700	.42	.12
Wetgeschiedenis (interne rechtsver- gelijking)	A (Pr)	.11	.00	F(1,13) =.650	.44	.12
	B (Pr)	.00	.00	-	-	-
	C (Sr)	.00	.00	-	-	-
Wetgeschiedenis (externe rechtsver- gelijking)	A (Pr)	.00	.17	F(1,13) =1.560	.23	.21
	B (Pr)	.00†	.25†	F(1,13) =3.178	.10	.38
	C (Sr)	.00	.00	-	-	-
Jurisprudentie (binnen rechtsge- bied)	A (Pr)	1.56	2.33	F(1,13) =1.613	.23	.22
	B (Pr)	2.25	2.53	F(1,13) =.267	.61	.08
	C (Sr)	1.25*	2.58*	F(1,14) =6.400	.02	.65
Jurisprudentie (interne rechtsver- gelijking)	A (Pr)	.44	.17	F(1,13) =.399	.54	.09
	B (Pr)	.55	.25	F(1,13) =.190	.67	.07
	C (Sr)	.00	.00	-	-	-
Jurisprudentie (externe rechtsver- gelijking)	A (Pr)	.00	.17	F(1,13) =1.560	.23	.21
	B (Pr)	.00†	.50†	F(1,13) =3.178	.10	.38
	C (Sr)	.00	.58	F(1,14) =.683	.42	.12
Literatuur (binnen rechtsgebied)	A (Pr)	1.00*	1.83*	F(1,13) =6.724	.02	.67

	B (Pr)	2.00	2.00	F(1,13) =.000	1.00	.05
	C (Sr)	.75	1.50	F(1,14) =1.500	.24	.21
Literatuur (interne rechtsvergelijking)	A (Pr)	.00	.00	-	-	-
	B (Pr)	.36	.50	F(1,13) =.074	.79	.06
	C (Sr)	.25	.08	F(1,14) =.700	.42	.12
Literatuur (externe rechtsvergelijking)	A (Pr)	.22	.33	F(1,13) =.118	.74	.06
	B (Pr)	.18	.75	F(1,13) =1.185	.30	.17
	C (Sr)	.00	.33	F(1,14) =.538	.48	.11
Literatuur (overig)	A (Pr)	.00	.00	-	-	-
	B (Pr)	.00	.00	-	-	-
	C (Sr)	.00	.17	F(1,14) =.318	.58	.08
Andere bronnen	A (Pr)	.00	.00	-	-	-
	B (Pr)	.00	.00	-	-	-
	C (Sr)	.00	.00	-	-	-
Gebrek aan bronnen	A (Pr)	3.00	2.00	F(1,13) =2.455	.15	.29
	B (Pr)	1.91	1.50	F(1,13) =.807	.39	.13
	C (Sr)	3.00***	1.33***	F(1,14) =25.000	.000	.996
Correct bronnengebruik	A (Pr)	3.50	4.50	F(1,13) =1.950	.19	.25
	B (Pr)	4.00	4.00	-	-	-
	C (Sr)	3.67	3.92	F(1,14) =1.232	.29	.18
Beschrijven/interpreteren (voldoende uitgebreid)	A (Pr)	-	-	-	-	-
	B (Pr)	2.30	3.00	F(1,12) =1.043	.33	.16
	C (Sr)	3.00	3.42	F(1,14) =.318	.58	.08
Beschrijven/interpreteren (voldoende diepgaand)	A (Pr)	-	-	-	-	-
	B (Pr)	2.50	3.00	F(1,12) =1.319	.27	.19
	C (Sr)	2.00	2.83	F(1,14) =2.500	.14	.31
Verduidelijken (voldoende uitgebreid)	A (Pr)	-	-	-	-	-

	B (Pr)	2.00*	3.50*	F(1,13) =7.800	<.02	.73
	C (Sr)	2.00*	3.58*	F(1,14) =7.059	<.02	.70
Verduidelijken (voldoende diep- gaand)	A (Pr)	-	-	-	-	-
	B (Pr)	2.36*	3.25*	F(1,13) =5.658	.03	.60
	C (Sr)	1.25***	3.50***	F(1,14) =36.98	.000	1.0
Ordenen (voldoen- de uitgebreid)	A (Pr)	-	-	-	-	-
	B (Pr)	2.45	3.00	F(1,13) =.891	.36	.14
	C (Sr)	.25*	2.50*	F(1,14) =6.697	.02	.67
Ordenen (voldoen- de diepgaand)	A (Pr)	-	-	-	-	-
	B (Pr)	2.55	2.50	F(1,13) =.007	.94	.05
	C (Sr)	.25*	2.33*	F(1,14) =7.172	<.02	.70
Doorrekenen (vol- doende uitgebreid)	A (Pr)	-	-	-	-	-
	B (Pr)	2.18	3.25	F(1,13) =6.813	.02	.68
	C (Sr)	.50	1.83	F(1,14) =2.286	.15	.29
Doorrekenen (vol- doende diepgaand)	A (Pr)	-	-	-	-	-
	B (Pr)	2.27	2.75	F(1,13) =1.761	.21	.23
	C (Sr)	.50	1.58	F(1,14) =2.061	.17	.27
Toetsen (overig) (voldoende uitge- breid)	A (Pr)	-	-	-	-	-
	B (Pr)	.73	1.33	F(1,12) =.337	.57	.08
	C (Sr)	.50	1.75	F(1,14) =2.100	.17	.27
Toetsen (overig) (voldoende diep- gaand)	A (Pr)	-	-	-	-	-
	B (Pr)	.27†	1.25†	F(1,13) =3.332	.09	.39
	C (Sr)	.50	1.67	F(1,14) =2.066	.17	.27
Essentiële vra- gen/aspecten gedis- tilleerd	A (Pr)	3.11†	4.00†	F(1,13) =4.160	.06	.47
	B (Pr)	3.27	3.75	F(1,13) =1.253	.28	.18

	C (Sr)	2.00***	4.00***	F(1,14) =28.00 0	.000	.998
Uitspraak nieuw	A (Pr)	2.13	2.50	F(1,13) =.402	.54	.09
	B (Pr)	2.27†	3.25†	F(1,13) =4.078	<.07	.46
	C (Sr)	3.00	2.50	F(1,13) =.553	.47	.11
Annotatie nieuw	A (Pr)	1.00	1.00	-	-	-
	B (Pr)	-	1.50	-	-	-
	C (Sr)	-	3.00	-	-	-
Reikwijdte uit- spraak	A (Pr)	2.00	2.17	F(1,13) =1.560	.23	.21
	B (Pr)	2.27*	3.25*	F(1,13) =7.385	.02	.71
	C (Sr)	2.75	2.50	F(1,14) =.269	.61	.08
Reikwijdte annota- tie	A (Pr)	2.00	2.00	-	-	-
	B (Pr)	2.00	-	-	-	-
	C (Sr)	-	2.50	-	-	-
Feitelijke informa- tie	A (Pr)	.00	.33	F(1,13) =1.560	.23	.21
	B (Pr)	.09	.00	F(1,13) =.347	.57	.09
	C (Sr)	.00	.33	F(1,13) =.700	.42	.12
Signaleren vraag- /knelpunten	A (Pr)	.88	1.17	F(1,13) =2.049	.18	.26
	B (Pr)	1.73†	2.50†	F(1,13) =4.394	<.06	.49
	C (Sr)	1.50*	2.92*	F(1,14) =7.073	<.02	.70
Vraag-/knelpunten door anderen gesig- naleerd	A (Pr)	-	3.00	-	-	-
	B (Pr)	2.17	1.67	F(1,7)= .467	.52	.09
	C (Sr)	4.00	3.00	F(1,11) =1.330	.27	.18
Constructieve bijdrage aan oplos- sen vraag-/knelpunt	A (Pr)	-	1.00	-	-	-
	B (Pr)	3.43	3.25	F(1,9)= .086	.78	.06
	C (Sr)	1.50	2.36	F(1,11) =1.971	.19	.25
Onderbouwd stand- punt	A (Pr)	-	-	-	-	-
	B (Pr)	2.14*	3.00*	F(1,9)= 5.891	.04	.58
	C (Sr)	1.00*	2.89*	F(1,8)= 5.255	.05	.52

Bespreken stand- punten/voorstellen andere auteurs	A (Pr)	1.00	1.20	F(1,12) =1.929	.19	.25
	B (Pr)	1.45	2.25	F(1,13) =2.546	.14	.32
	C (Sr)	1.25	1.60	F(1,12) =4.59	.51	.10
Weerleggen stand- punten andere auteurs	A (Pr)	-	1.33	-	-	-
	B (Pr)	2.00	2.67	F(1,6)= .750	.42	.11
	C (Sr)	.50	1.67	F(1,3)= 4.200	.13	.30
Praktische uitspraak	A (Pr)	1.89	2.50	F(1,13) =2.736	.12	.34
	B (Pr)	2.55	3.25	F(1,13) =2.532	.14	.31
	C (Sr)	2.50	3.00	F(1,13) =.867	.37	.14
Praktische annotatie	A (Pr)	1.11	1.00	F(1,13) =.650	.44	.12
	B (Pr)	2.18*	3.33*	F(1,12) =5.951	.03	.61
	C (Sr)	1.50**	2.60**	F(1,12) =12.20	.004	.89

Noot: † p < .10. * p < .05. ** p < .01. *** p < .001.

Globale samenvattingen van de gecodeerde annotaties

1 Bestuursrecht

Annotatie nr. 1

[Niet gebruikt. De auteur van de annotatie kwam overeen met een van de beoordelaars.]

Annotatie nr. 2, arrest Europees Hof van Justitie (2 uit 2 goed/uitstekend)

De noot, die ongeveer 3,5 pagina's beslaat, hoort bij een arrest dat gaat over een beslissing van het Europese Hof van Justitie over de toepassing van bepalingen uit het EG-verdrag. De vraag die aan de orde is, is of de bepalingen ambtshalve moeten worden toegepast. De noot begint te analyseren of de uitspraak overeenkomt met een eerder genomen beslissing, waar het Hof ook naar verwijst. Vervolgens stelt de annotator in ongeveer een halve pagina en aan de hand van een beschrijving van het geval vast dat het in het onderhavige geval ging om het ambtshalve aanvullen van gronden en niet (noodzakelijk) om ambtshalve toetsing. De conclusie is dat de uitspraak geen ambtshalve toetsing impliceert en dat hoe dit ook zij, met het oordeel nog niet is gezegd hoe ver die ambtshalve toepassing dient te gaan. De annotator vervolgt met het in een aantal regels wijzen op het belang van de uitspraak. Daarbij geeft hij de belangrijkste overwegingen van het Hof, wat ongeveer een halve pagina in beslag neemt. In de laatste regels geeft de annotator aan in welke richting het rechtsgebied zich op dit specifieke punt zal ontwikkelen.

Annotatie nr. 3, uitspraak Afdeling Bestuursrechtspraak (2 uit 2 goed/uitstekend)

De kwestie die in de uitspraak aan de orde is, gaat over een ingetrokken subsidie. De annotatie begint met een beschrijving van de feiten en het geldende recht dat op de kwestie van toepassing is (circa anderhalve pagina). In de volgende twee pagina's wijst de annotator op een praktische implicatie van de uitspraak. De annotator gaat hierbij achtereenvolgens in op hoe vergelijkbare beslissingen in andere rechtsgebieden van het Nederlandse recht uitpakken. Op basis van die vergelijkingen komt de annotator tot de conclusie dat het voor de hand had gelegen dat de bestuursrechter in dit geval een uitzondering had aangenomen op het beginsel dat aan de orde is in de uitspraak. Ten slotte wordt een motiveringsgebrek opgemerkt en een verklaring gegeven waarom zo geoordeeld is in deze zaak.

Annotatie nr. 4, uitspraak Afdeling Bestuursrechtspraak (1 uit 2 goed/uitstekend)

Deze ruim drie pagina's tellende noot opent met een beschrijving van het geldende recht (één derde pagina). Hierna volgt een samenvatting van de uitspraak. Dit onderdeel beslaat ruim een halve pagina. Vervolgens wordt een motiveringsgebrek opgemerkt en een argument tegen de beslissing gegeven. Daarna wordt in ruim anderhalve pagina uitgebreid kritiek op de beslissing gegeven, met in enkele regels een vergelijking met eerdere jurisprudentie. De annotator gaat daarbij in op het gevaar van het gebruik van bepaalde interpretatietechnieken en op het nalaten een begrip te definiëren dat in de uitspraak aan de orde is. De noot besluit met kritiek op een, in deze uitspraak, te weinig kritische houding van de rechter.

Annotatie nr. 5, uitspraak Afdeling Bestuursrechtspraak (1 uit 2 goed/uitstekend)

Eerst wordt de voorliggende uitspraak samengevat en het geldend recht beschreven in ruim één pagina (van de 2,5 in totaal). In een tiental regels toetst de annotator vervolgens of de uitspraak voldoet aan het doel van de regeling. Hierna, in de laatste pagina, geeft de annotator aan het niet met de uitspraak eens te zijn en doet een voorstel dat afwijkt van voorstellen van andere auteurs.

Annotatie nr. 6, uitspraak Afdeling Bestuursrechtspraak (2 uit 2 goed/uitstekend)

In de drie pagina's tellende annotatie begint de annotator met ruim anderhalve pagina waarin hij de uitspraak samenvat en het geldend recht bespreekt. De tweede helft van de noot bestaat uit een beschrijving van een opvatting waarom het bestuursrecht op dit punt niet aansluit bij maatschappelijke verwachtingen. De annotatie vervolgt met de vraag of zaken zoals aan de orde in de uitspraak niet eenvoudiger kunnen worden afgewikkeld, in die zin dat sneller een uitkomst in het materiële geschil kan worden verkregen. De annotator denkt dat het inderdaad mogelijk is om sneller tot een oplossing te komen, maar dat zich een knelpunt kan voordoen, welk knelpunt hij noemt en uitwerkt.

Annotatie nr. 7

[Niet gebruikt. Geen commentaar bij een uitspraak van een rechter.]

Annotatie nr. 8, uitspraak Afdeling Bestuursrechtspraak (1 uit 2 goed/uitstekend)

De noot begint met en bestaat voor het grootste deel uit een beschrijving van het geldende recht (ruim de helft van anderhalve pagina in totaal). De annotator

merkt in dat verband op dat wat de wet voorschrijft afwijkt van wat rechters doorgaans doen. Hierna volgt een samenvatting van de zaak (iets minder dan een halve pagina) met direct daarop een evaluatie van de beslissing aan de hand van de gangbare praktijk. Er wordt beargumenteerd dat een bepaalde procedure geen meerwaarde heeft omdat rechters in die procedure niet doen wat van hen wordt verlangd.

Annotatie nr. 9, uitspraak Afdeling Bestuursrechtspraak (2 uit 2 goed/uitstekend)

Deze bijna vier pagina's lange noot bestaat voor ongeveer de helft uit de beschrijving van het geldend recht en de samenvatting van de onderhavige zaak. Verder wordt regelmatig het belang van de uitspraak geduid alsook de implicaties van de uitspraak, steeds in een aantal regels. De annotator toetst de uitspraak verder (in enkele regels) aan een verdrag en constateert dat het criterium dat de Afdeling gebruikt een verdragsrechtelijke toets doorstaat. Een vergelijking met eerdere jurisprudentie levert op dat de rechter een nieuwe koers is gaan varen, maar dat deze koers in haar uitwerking niet helder is. Ten slotte uit de annotator in de laatste alinea (ongeveer tien regels) kritiek op voorstellen die gedaan zijn door Kamerleden.

Annotatie nr. 10, uitspraak Afdeling Bestuursrechtspraak (2 uit 2 goed/uitstekend)

De beschrijving van het geldende recht en een samenvatting van de onderhavige zaak beslaan ruim één pagina van de bijna twee pagina's lange noot. Voor het overige bestaat de noot hoofdzakelijk uit een beschrijving van kritiek die in de literatuur is geuit op de lijn die in de rechtspraak wordt gevolgd. De noot besluit met de verwachting dat de Afdeling op den duur een andere koers zal varen.

Annotatie nr. 11, uitspraak Afdeling Bestuursrechtspraak (1 uit 2 goed/uitstekend)

De noot van één pagina lang bestaat op twee regels na uit een samenvatting van de onderhavige zaak en daarnaast – voor een kleiner deel – uit beschrijving van het geldend recht. De andere twee regels behelzen de opmerking dat het geval een voorbeeld is van wat een plaats zou moeten krijgen in een nieuwe algemene regeling.

Annotatie nr. 12, uitspraak Afdeling Bestuursrechtspraak (1 uit 2 goed/uitstekend)

Een derde van deze korte (nog geen halve pagina) noot bestaat uit een samenvatting van de onderhavige zaak, gevolgd door een beschrijving van het geldende

recht (een derde deel van de noot). De annotator toetst, tot slot, of de beslissing van de rechter voldoet aan het doel van de regeling.

Annotatie nr. 13, arrest Hoge Raad (0 uit 2 goed/uitstekend)

De noot is ruim 11 pagina's lang en begint met 5,5 pagina's waarin sprake is van een combinatie tussen een samenvatting van de uitspraak en een beschrijving van het geldende recht, waar later in de noot nog 1,5 pagina aan wordt toegevoegd. Hierna volgt 1,5 pagina waarin de discussie omtrent het onderwerp in de literatuur wordt besproken. Hierin wordt verslag gedaan van een discussie die in de literatuur gaande is over hoe de problematiek zou moeten worden geregeld. Ook wordt één pagina besteed aan het vergelijken van de uitspraak met eerdere jurisprudentie. De vergelijking resulteert in de constatering dat de uitspraak voor een deel aansluit bij eerdere rechtspraak en dat dit voor een deel niet op te maken is.¹⁶¹ Aan het slot noemt de annotator argumenten die volgens hem een rol hebben gespeeld bij de beslissing maar die niet in de uitspraak worden genoemd (ongeveer een halve pagina).

Annotatie nr. 14

[Geen annotatie.]

Annotatie nr. 15, uitspraak Centrale Raad van Beroep (0 uit 2 goed/uitstekend)

Het grootste deel van deze noot bestaat uit een samenvatting van de onderhavige zaak (één pagina van de 1,5 pagina tellende noot). Onder verwijzing naar de verwachtingen van rechtzoekenden stelt de annotator dat de uitspraak niet aansluit bij hoe de bestuursrechter te werk gaat. De annotator geeft aan van mening te zijn dat de rechter in dit geval de verkeerde keuze heeft gemaakt en heeft op die manier kritiek op de uitspraak.

Annotatie nr. 16, arrest Hoge Raad (0 uit 2 goed/uitstekend)

Deze relatief korte noot bestaat uit een samenvatting van de uitspraak en uit een weergave van de regels die op de problematiek van toepassing zijn.

Annotatie nr. 17, uitspraak College van Beroep voor het bedrijfsleven (0 uit 2 goed/uitstekend)

Deze twee pagina's tellende noot bestaat uit een samenvatting van de onderhavige uitspraak en uit het geldende recht dat op het geval van toepassing is.

¹⁶¹ Uit de vergelijking volgen bijvoorbeeld geen nieuwe gezichtspunten of criteria die bruikbaar kunnen zijn bij het beslechten van toekomstige gevallen.

Annotatie nr. 18, uitspraak Afdeling Bestuursrechtspraak (1 uit 2 goed/uitstekend)

De noot bestaat voor twee derde uit een samenvatting en beschrijving van de uitspraak en van het geldende recht, waarbij naar relevante literatuur wordt verwezen. Na het aanstippen van een onduidelijkheid volgt kritiek op de toepassing van de regel door de rechter. Onder verwijzing naar eerdere uitspraken wordt gesteld dat de rechter minder marginaal toetst dan in eerdere jurisprudentie.

2 Privaatrecht

Annotatie nr. 1, arrest Europees Hof van Justitie (0 uit 3 goed/uitstekend)

De annotatie bestaat voor een groot deel uit een weergave van de feiten, de uitspraak en van het geldende recht dat erop van toepassing is. In dit verband wijst de annotator tevens op het belang van de uitspraak, dat er uit bestaat dat de rechter bepaalde eisen nogmaals duidelijk heeft opgenomen in de uitspraak. Aan het slot van de noot wordt een vergelijking gemaakt met een eerdere uitspraak, die bestaat uit de constatering dat de uitspraak in kwestie overeenstemt met die betreffende eerdere uitspraak.

Annotatie nr. 2, arrest Hoge Raad (3 uit 3 goed/uitstekend)

De annotatie valt uiteen in een beschrijvend deel en in een meer evaluerend deel. Het beschrijvende deel bestaat hoofdzakelijk uit een weergave van het geldende recht, waarbij meerdere malen (vijf keer) wordt vergeleken met eerdere uitspraken, dat wil zeggen, met een enkele zin wordt opgemerkt welke onderdelen uit de uitspraak overeenkomen met eerdere jurisprudentie en welke elementen nieuw of anders zijn. Daarbij geeft de annotator literatuurverwijzingen als hij een vraag opwerpt die uitvoerig naar aanleiding van eerdere uitspraken is besproken in de juridische literatuur. Verder constateert de annotator een onduidelijkheid in de motivering, die hij opheldert door een passage uit de uitspraak te herschrijven.

In het evaluatieve deel van de annotatie loopt de annotator vier vragen na die bij de annotator zijn gerezen na het lezen van de uitspraak. De bespreking van deze vragen beslaat een substantieel deel van de annotatie (circa twee van de vier pagina's). Eén vraag komt neer op het bestaan van een onduidelijke passage, die de annotator vervolgens tracht te verduidelijken. Bij een andere vraag vindt de beantwoording plaats door te verwijzen naar een publicatie over dit onderwerp. De twee andere vragen worden besproken aan de hand van nieuwe, zelfbedachte voorbeelden. Met deze casus gaat de annotator onder andere na of het resultaat dat met de toepassing van het criterium dat de rechter heeft gehanteerd als redelijk kan worden beschouwd. Dat criterium leidt hij af uit de uitspraak, waarmee de annotator een bijdrage levert aan de interpretatie en toepas-

sing van het geldende recht op dit punt. De annotator besluit met het aanbrengen van een nadere nuancering van de regel zoals die uit het arrest kan worden afgeleid, mede onder verwijzing naar een andere uitspraak, en zet dit alternatief af tegen de beslissing in de besproken uitspraak. Daarbij komt hij tot de conclusie dat de ruimte die het alternatief biedt overeenkomt met de ruimte zoals die in een eerdere uitspraak werd geboden, maar waarin een andere grondslag voor die beoordelingsvrijheid was gekozen.

Annotatie nr. 3, arrest Hoge Raad (0 uit 3 goed/uitstekend)

Na de constatering dat de uitspraak waarbij de annotatie is geschreven, een nadere uitwerking is van een eerder arrest, geeft de annotator een samenvatting van de feiten en de uitspraak. De uitspraak bevat enkele voorbeelden van omstandigheden die voor de beoordeling van een bepaald type geschillen of vragen relevant zijn. De annotator brengt deze voorbeelden terug tot een drietal categorieën en licht deze categorieën toe. De noot vervolgt met een bespreking van de implicaties van de uitspraak, in die zin dat de annotator aan de hand van voorbeelden en de uitspraak zelf bespreekt wat de rechter heeft bedoeld. De annotator besluit de noot met verwijzingen naar uitspraken die qua feitenconstellatie verschillen van de uitspraak waar de annotatie bij is geschreven, maar waar zich wel dezelfde problematiek voordoet.

Annotatie nr. 4, arrest Hoge Raad (3 uit 3 goed/uitstekend)

Na een weergave van de feiten en het oordeel van de Hoge Raad, bespreekt de annotator het geldende recht ten aanzien van het onderwerp dat in de uitspraak centraal staat. Hij verwijst daarbij naar buitenlands recht en gaat aan de hand daarvan in op de oorsprong en het ontstaan van het leerstuk en geeft daarbij gemotiveerd aan waarom een redenering uit de wetsgeschiedenis en literatuur niet overtuigt. De annotatie vervolgt met het ontwikkelen van een criterium aan de hand waarvan een beslissing kan worden genomen in gevallen zoals in de uitspraak. Het ontwikkelen van dit criterium gebeurt door een criterium te bekritisseren dat in de juridische, privaatrechtelijke literatuur is genoemd en door dit eerder aangedragen criterium te toetsen aan het beginsel van rechtszekerheid, welk beginsel in het betreffende rechtsgebied een belangrijke rol speelt. Daarna toetst de annotator de uitspraak aan het zojuist ontwikkelde criterium en komt tot de conclusie dat de rechter dit criterium heeft toegepast. Daarmee toetst de annotator of de beslissing beantwoordt aan de beginselen die ten grondslag liggen aan het betreffende rechtsgebied of leerstuk, zij het dat de toets een impliciete is. Vervolgens verkent de annotator de reikwijdte en betekenis van de uitspraak door na te gaan welke feiten anders dienen te zijn willen deze resulteren in een andere uitkomst. De noot sluit af met de onderbouwde stelling dat de procespartijen, en partijen meer in het algemeen, beter een (andere) juridische

constructie of grondslag voor hun afspraken hadden kunnen kiezen dan de constructie die aan de orde is in de betreffende uitspraak.

Annotatie nr. 5, arrest Hoge Raad (1 uit 3 goed/uitstekend)

Onder verwijzing naar het geldende recht dat op gevallen als deze van toepassing is, begint de annotatie met een bespreking van de feiten en het oordeel van de rechter. De annotator stipt daarbij het belang van de uitspraak aan. Dat belang schuilt volgens de annotator in de bredere toepasbaarheid van de invulling van het criterium dat in de uitspraak door de rechter wordt ingevuld dan wel aangescherpt. Andere privaatrechtelijke leerstukken kennen een vergelijkbaar criterium, wat de vraag oproept of de uitspraak mag worden doorgetrokken naar die andere leerstukken. In de praktijk wordt dat volgens de annotator bestreden, maar de annotator stelt hiervan dat vooral de verschillen worden benadrukt in plaats van de overeenkomsten.¹⁶² De annotator bespreekt de verschillende leerstukken in een aantal pagina's en gaat in op de vraag op welke punten de besproken uitspraak relevant is voor die andere leerstukken. De annotator concludeert dat de uitspraak relevant is voor ten minste één ander leerstuk binnen het privaatrecht.

Annotatie nr. 6, arrest Hoge Raad (1 uit 3 goed/uitstekend, 1 uit 3 matig/slecht, 1 uit 3 matig/goed)

De noot bestaat voor een groot deel uit een weergave van de feiten en het oordeel van de rechter. Deze weergave gaat gepaard met een beschrijving van het geldende recht, hetgeen achtergrondinformatie geeft over de beslissing van de rechter. Op een aantal plaatsen behandelt de annotator andere aspecten. De eerste keer stelt en motiveert hij in een tiental regels dat hij het eens is met de beslissing van de rechter, ondanks dat het wettelijk systeem hiervoor dient te worden opgerekt. De tweede keer stelt de annotator vast dat het oordeel van de rechter niet eenvoudig in het wettelijke systeem te plaatsen is en geeft hij een interpretatie van de uitspraak die zich laat verdragen met dat wettelijke systeem. De derde keer waarin de annotator verder gaat dan het weergeven van de feiten, uitspraak en de relevante regels, is als hij in enkele regels de implicaties bespreekt, waarbij hij vervolgens refereert aan eerdere uitspraken. Uit die vergelijking concludeert de annotator dat deze uitspraak in lijn is met die andere uitspraken en bovendien qua geval ervan afwijkt. Verder stelt de annotator ook op een tweetal andere plaatsen dat hij het eens is met een overweging van de rechter, maar verder dan deze constatering gaan de opmerkingen niet. Datzelfde

¹⁶² De beoordelaar die deze annotatie niet als goed/uitstekend heeft aangemerkt, voerde als reden aan dat de annotator de uitspraak niet had mogen doortrekken naar de andere privaatrechtelijke leerstukken zoals die in de annotatie aan de orde komen.

geldt voor de opmerking van de annotator waarbij hij aangeeft onder welke omstandigheid wel aan een bepaald criterium zou zijn voldaan.

Annotatie nr. 7, arrest Hoge Raad (1 uit 3 goed/uitstekend)

De noot begint met het schetsen van het belang van de uitspraak en werpt vervolgens een aantal vragen op waar de uitspraak een antwoord op geeft. De annotator loopt deze vragen achtereenvolgens na, daarbij verwijzend naar feiten, overwegingen uit de uitspraak en naar het geldende recht. Dit beslaat het grootste deel van de noot. Op één plaats stipt de annotator in een aantal zinnen de implicaties van de uitspraak aan. Verder maakt de annotator op een zeker moment een vergelijking met een uitspraak waar ook in de uitspraak aan wordt gerefereerd, maar deze vergelijking gaat niet verder dan de opmerking dat de besproken uitspraak ziet op een ander (type) geval dan de eerder gewezen uitspraak.

Annotatie nr. 8, arrest Hoge Raad (2 uit 3 goed/uitstekend)

Het eerste deel van de annotatie bestaat uit een weergave van de feiten, het oordeel van de rechter en het geldende recht dat op het geval van toepassing is, waaronder ook het verwijzen naar literatuur en jurisprudentie. De noot vervolgt met het schetsen van de implicaties van de uitspraak voor de interpretatie en toepassing van het geldende recht in toekomstige gevallen, wat ruim een alinea beslaat. Na kort een verklaring te hebben gegeven voor het oordeel van de rechter en een motiveringsgebrek te hebben opgemerkt, vervolgt de noot met een vergelijking met eerdere uitspraken. Deze vergelijking resulteert, kort samengevat, in de conclusie dat de uitspraak qua motivering afwijkt van eerdere uitspraken, maar dat dit verschil niet tot gevolg heeft dat ook de uitkomst een andere zou zijn. Op grond hiervan verwacht de annotator niet dat de uitspraak zal leiden tot een toename van de onzekerheid omtrent het leerstuk dat in de annotatie centraal staat.

Annotatie nr. 9, arrest Hoge Raad (0 uit 3 goed/uitstekend)

De noot bestaat voor het grootste deel uit een beschrijving van het wettelijke systeem zoals dat van toepassing is op gevallen zoals aan de orde in de uitspraak. Aan het slot van de annotatie wordt opgemerkt dat de uitspraak in kwestie past in de lijn zoals de Hoge Raad die eerder al had ingezet. Deze opmerking wordt gevolgd door een korte weergave van het oordeel van de rechter. Tot slot wijst de annotator in enkele regels op de gevolgen van de uitspraak indien deze wordt doorgetrokken naar een ander gevaltype, waarbij hij opmerkt dat eerdere, lage rechtspraak niet in lijn is met de gevolgtrekking die hij schetst.

Annotatie nr. 10, twee arresten Hoge Raad (2 uit 3 goed/uitstekend)

De annotatie, die hoort bij twee arresten, begint met het schetsen van het geldende recht, waarbij de annotator (drie keer) kort aanstipt waar de besproken uitspraak aansluit dan wel afwijkt van eerder verschenen jurisprudentie. De noot vervolgt met een alinea over de implicaties van de uitspraak voor de praktijk van bewijsvoering. Hierna gaat de noot verder met een weergave van de overwegingen van de rechter, afgewisseld met (twee) passages waarin de annotator gezichtspunten uit de twee uitspraken distilleert, deze samenbrengt en constateert dat de gezichtspunten overeenkomen met de gezichtspunten zoals die in de literatuur naar voren komen.

Annotatie nr. 11, arrest Hoge Raad (1 uit 3 goed/uitstekend)

Deze noot bestaat voor het grootste deel uit een weergave van de feiten, de uitspraak en het geldende recht dat daarop van toepassing is. Op een zeker moment stelt de annotator ten aanzien van een overweging dat hij het hiermee eens is. Hij onderbouwt dit door te laten zien dat de overweging van de rechter aansluit bij het doel van de wettelijke regeling. Op een tweetal plaatsen bespreekt de annotator de implicaties van de uitspraak. De eerste keer bespreekt hij wat de rechter zou doen indien zich een bepaalde situatie zou voordoen. De tweede keer gaat de annotator in op de gevolgen van de uitspraak voor procespartijen, meer specifiek op wie wat dient aan te tonen. Verder wordt met een enkel woord geconstateerd dat de uitspraak in lijn is met een andere uitspraak.¹⁶³

Annotatie nr. 12, arrest Hoge Raad (2 uit 3 goed/uitstekend)

Deze noot bleek moeilijk te coderen, wat het gevolg is van typen passages die in elkaar overlopen. Zo loopt een bespreking van de uitspraak, het beschrijven van het wettelijke systeem en het evalueren van de uitspraak op verschillende punten in de noot door elkaar. Het eerste deel van de annotatie bevat een beschrijving van het wettelijke systeem dat op het besproken geval van toepassing is en zoals dat door de rechter is toegepast. Dit eerste deel kan dan ook worden gezien als een verduidelijking of nadere toelichting van de uitspraak. De noot vervolgt met een weergave van de feiten en het oordeel van de rechter, waarbij de annotator op drie plekken aangeeft het eens te zijn met de overwegingen en conclusies van de rechter. Verder evalueert de annotator op één punt de uitspraak door te bespreken wat de consequenties zouden zijn geweest als de rechter anders zou hebben geoordeeld. Daarnaast maakt de noot op één punt (kort) een vergelijking met een eerdere uitspraak, die bestaat uit de opmerking dat een eerdere uitspraak geen aanleiding geeft om anders te oordelen dan de rechter heeft gedaan in de besproken uitspraak. Vervolgens spitst de noot zich toe op de vraag welke feiten

¹⁶³ Ik heb dit gecodeerd als een vergelijking van de uitspraak met een eerdere uitspraak, maar daarbij moet worden aangetekend dat de vergelijking zowel summier als mager is.

en omstandigheden relevant zijn voor het invullen van bepaalde regels. Daartoe wordt het oordeel van de rechter besproken, alsook de wettelijke systematiek, gevolgd door een bespreking van de implicaties van het oordeel. De noot sluit af met een conclusie die voornamelijk bestaat uit een samenvatting van de uitspraak en het schetsen van achtergronden.

Annotatie nr. 13, vonnis rechtbank (0 uit 3 goed/uitstekend)

De annotatie bestaat overwegend uit een weergave van de feiten uit de uitspraak en het oordeel van de rechter. Op een tweetal punten wordt met enkele zinnen verwezen naar het geldende recht (rechtspraak).

Annotatie nr. 14, arrest Hoge Raad (1 uit 3 goed/uitstekend)

Na een beschrijving van het geldende recht en de belangrijkste feiten en overwegingen uit de uitspraak, vergelijkt de annotator de besproken uitspraak met eerdere beslissingen. Daartoe vat hij die eerdere beslissingen samen, beschrijft hij op welke onderdelen de casus verschilt van die uit eerdere gevallen en constateert de annotator dat het criterium dat de rechter in de besproken uitspraak heeft aangelegd niet verschilt van eerdere uitspraken. Omdat de uitspraak een deelgebied betreft van een meer algemeen leerstuk, vervolgt de noot met een bespreking van de relatie tussen de besproken uitspraak en het meer algemene leerstuk. Na de kernuitspraken behorend bij het algemene leerstuk te hebben besproken, merkt de annotator in een aantal regels op dat het criterium dat in de besproken uitspraak wordt aangelegd, ook aan de orde is in een van de standaardarresten. Niet besproken wordt wat de gevolgen of implicaties daarvan zijn. Vervolgens grijpt de annotator terug op een eerdere uitspraak om te beschrijven hoe het gebruikte criterium moet worden toegepast en laat zien wat het besproken arrest hieraan toevoegt. Daarna gaat de noot in op een tweetal andere aspecten die in de uitspraak aan de orde zijn. Van het eerste aspect concludeert de annotator dat hij het eens is met wat de rechter hierover heeft gesteld, zich daarbij beroepend op een verdrag waarin hetzelfde is bepaald. Ten aanzien van het tweede aspect constateert de annotator dat de rechter de voorwaarden herhaalt zoals die eerder al door de hoogste rechter waren geformuleerd. De noot besluit met de opmerking dat de rechter een niet al te formele opstelling heeft gekozen, wat de annotator toejuicht.

Annotatie nr. 15, vonnis rechtbank (0 uit 2 goed/uitstekend)

De annotatie bestaat voor het grootste deel uit een weergave van de casus, het oordeel van de rechter en het recht dat van toepassing is op de problematiek. Op drie plaatsen in de annotatie wijkt de annotatie hiervan af. Twee keer stelt en motiveert de annotator in enkele zinnen respectievelijk een alinea dat hij het niet eens is met wat de rechter heeft bepaald. Eén keer gebeurt dit onder verwijzing

naar een eerdere uitspraak. De derde keer dat de annotator meer doet dan het samenvatten van de casus en de uitspraak is wanneer hij ingaat op de (hypothetische) situatie dat een van de procespartijen een andere keuze zou hebben gemaakt.

Annotatie nr. 16

[Niet gebruikt. De publicatie werd door twee van de drie beoordelaars aangemerkt als een mogelijk artikel. De derde beoordelaar gaf bij nadere vragen ook aan twijfels te hebben.]

3 Strafrecht

Annotatie nr. 1, arrest Hoge Raad (2 uit 2 goed/uitstekend)

Ongeveer één derde van deze 3,5 pagina's tellende noot bestaat uit een samenvatting van de zaak en een beschrijving van geldend recht. De annotator gaat in op een door de verdachte aangevoerde innerlijke tegenstrijdigheid en heldert deze op. Ook geeft de annotator aan het begin van de noot het belang van de uitspraak aan. Een groot deel (1,5 pagina) van de noot is gereserveerd voor het weergeven van verschillende opvattingen in de literatuur over het centrale leerstuk in deze zaak en welke leer door de Hoge Raad is toegepast. Na het aangeven van de implicaties van de zaak sluit de noot af met meer beschrijving van het geldend recht en de literatuur hierover (ongeveer 0,5 pagina).

Annotatie nr. 2, arrest Hoge Raad (2 uit 2 goed/uitstekend)

De noot heeft een omvang van ongeveer drie pagina's lang en begint met (onderbouwde) kritiek op de uitspraak. Ook wordt een motiveringsgebrek opgemerkt en worden kort de implicaties van de uitspraak weergegeven. Een relatief klein deel van de noot bestaat uit een samenvatting van de uitspraak. De annotatie bestaat hoofdzakelijk (bijna twee pagina's) uit het nalopen van argumenten voor en tegen de keuze die in het Nederlandse recht ten aanzien van strafrechtelijke aansprakelijkheid is gemaakt. De annotator loopt daartoe verschillende gedachten af die de keuze verklaren (rechtvaardigen) en concludeert steeds dat deze redenen de keuze niet kunnen rechtvaardigen.

Annotatie nr. 3, arrest Hoge Raad (2 uit 2 goed/uitstekend)

De noot heeft een omvang van ongeveer twee pagina's en begint met een combinatie van het samenvatten van de uitspraak en het beschrijven van het geldende recht. Dit deel beslaat ongeveer de helft van de noot. Vervolgens wordt een mogelijke verklaring gegeven voor waarom de rechter zo beslist heeft (ongeveer een halve pagina). Hierna worden implicaties van de uitspraak besproken voor de toepassing van het leerstuk. In dit verband komt de annotator tot de conclusie

dat vraagtekens kunnen worden geplaatst bij de theoretische grondslag van de rechtsplicht die in de uitspraak is aangenomen; wat de waarde van de grondslag is indien deze louter afhankelijk is van de omstandigheden van het geval. Na deze constatering plaatst de annotator een tweetal opmerkingen, eentje over een mogelijke vergissing in de uitspraak, de andere over een passage in de tenlastellegging.

Annotatie nr. 4, arrest Hoge Raad (2 uit 2 goed/uitstekend)

Aan het begin van de noot wordt het belang van de uitspraak aangestipt en worden verschillende opvattingen uit de literatuur over dit onderwerp beschreven (ongeveer een halve pagina). Het grootste deel van de noot bestaat verder uit samenvattingen van de uitspraak afgewisseld met vergelijking van de uitspraak aan eerdere jurisprudentie (ongeveer driekwart pagina). Deze vergelijking dient ertoe vast te stellen welke benadering – in de literatuur worden meerdere mogelijke verklaringsmodellen gegeven – de jurisprudentie van de Hoge Raad verklaart. Op grond van de vergelijking concludeert de annotator dat de ene benadering een grotere verklaringskracht heeft dan de andere. Afgesloten wordt met een bespreking van de implicaties van de uitspraak.

Annotatie nr. 5, arrest Hoge Raad (2 uit 2 goed/uitstekend)

Deze annotatie, die 3,5 pagina's telt, heeft veel weg van annotatie nr. 4. De annotatie begint met een weergave van de casus en een beschrijving van het geldend recht, waaruit ongeveer één derde van de noot bestaat. Hierna volgt een vrij uitgebreide (ruim 0,5 pagina) beschouwing van verschillende theorieën over het centrale punt in deze uitspraak. Vier maal worden de implicaties besproken van de uitspraak, wat in totaal bijna één pagina beslaat. Vervolgens bespreekt de annotator een andere kwestie. Hij gaat in op de vraag of een bepaald aspect een rol heeft gespeeld in de uitspraak, en concludeert dat de passage die suggereert dat het aspect een rol zou kunnen hebben gespeeld, slechts voorbereidt op een andere overweging. De annotator besluit de annotatie door enkele beleidsmatige overwegingen aan de orde te stellen.

Annotatie nr. 6, arrest Hoge Raad (2 uit 2 goed/uitstekend)

De noot begint met het in enkele regels aanduiden van het belang van de uitspraak. Vervolgens wordt de casus en het geldende recht beschreven, waar in totaal ongeveer één pagina van de noot uit bestaat. Mede naar aanleiding van de conclusie van de A-G constateert de annotator in dit verband dat de opvattingen in de literatuur en rechtspraktijk uiteenlopen en dat de Hoge Raad in dit geval heeft gekozen voor een specifieke benadering. Nadat in enkele regels een onduidelijkheid in de uitspraak is verhelderd, wordt de casus en het geldende recht verder besproken. Daarbij worden, kort, drie maal implicaties van de uitspraak

aangestipt. Twee keer vindt een vergelijking plaats met eerdere jurisprudentie, waarbij wordt geconcludeerd dat de uitspraak van de Hoge Raad aansluit bij eerdere uitspraken. Afgesloten wordt met ruim een halve pagina over implicaties van de uitspraak, meer specifiek hoe een bepaald begrip moet worden opgevat en worden toegepast. In dit kader merkt de annotator op dat in de literatuur een uitholling van het begrip is verdedigd, maar dat uit de uitspraak in kwestie, in vergelijking met eerdere uitspraken, volgt dat van uitholling waarschijnlijk geen sprake is.

Annotatie nr. 7, arrest EHRM (2 uit 2 goed/uitstekend)

Deze vier pagina's tellende noot bestaat voor ongeveer drie pagina's uit beschrijving van het geldend recht en weergave van de casus. De annotator begint zijn noot met de vraag of de uitspraak van het EHRM in strijd is met de lijn die het zelf heeft uitgezet. Door een reeks uitspraken te vergelijken, tracht de annotator dit te verklaren. Dit leidt tot de conclusie dat de uitspraak op een aantal belangrijke punten verschilt van andere, eerdere uitspraken. Naar aanleiding hiervan concludeert de annotator dat de casuïstiek dwingt tot nuancering van de regels. Vervolgens gaat de annotator uitvoerig in op de verhouding tussen de rechtspraak van de Hoge Raad en die van het EHRM ten aanzien van het punt dat in de uitspraak centraal staat. Aan de hand van een bespreking van een hoeveelheid uitspraken betoogt de annotator dat het Hof en de Hoge Raad feitelijk hetzelfde criterium hanteren, maar dat de algemene regel van de Hoge Raad onder druk staat. Het bespreken van de implicaties van het arrest van het EHRM voor de Nederlandse rechtspraak neemt ongeveer twee derde pagina in beslag.

Annotatie nr. 8, arrest Hoge Raad (2 uit 2 goed/uitstekend)

De annotator sluit zich in de noot aan bij wat in de uitspraak is bepaald, maar wil juist ingaan op de afwijkende mening van de A-G. De noot bestaat voor een klein deel uit een samenvatting van de uitspraak (ongeveer een halve pagina van de 2,5) en een ongeveer even groot deel bestaat uit de beschrijving van geldend recht. Nog eens een halve pagina wordt besteed aan het onderbouwd aangeven dat de annotator het met de uitspraak eens is (driekwart pagina). De annotator geeft daarbij aan dat in de literatuur ook anders wordt gedacht over hoe om te gaan met kwesties zoals in de uitspraak, welke opvattingen worden beschreven. Een relatief groot deel wordt besteed aan het bespreken van implicaties van de uitspraak (en de lijn die de Hoge Raad al langer volgt).

Vervolgens gaat de annotator in op de taak van de rechter, en concreetiseert deze, onder verwijzing naar literatuur en jurisprudentie. Uit die analyse leidt de annotator af wat de rechter had moeten doen en constateert dat dit ook is wat de rechter heeft gedaan in de besproken uitspraak.

Annotatie nr. 9, arrest Hoge Raad (0 uit 2 goed/uitstekend)

Iets minder dan de helft van deze (anderhalve pagina tellende) noot bestaat uit een samenvatting van de uitspraak. Tweemaal wordt in een aantal zinnen vergeleken met een uitspraak van het EHRM. Uit deze vergelijking wordt de conclusie getrokken dat de besproken zaak verschilt van het EHRM-geval. Voorts is een alinea besteed aan het schetsen van de problemen die zich geregeld voor zouden kunnen doen rondom deze problematiek. Voor het overige bestaat de noot uit een beschrijving van geldend recht.

Annotatie nr. 10, arrest Hoge Raad (1 uit 2 goed/uitstekend)

Deze noot is slechts één pagina lang en ongeveer een derde daarvan bestaat uit een samenvatting van de uitspraak. Daarnaast bestaat iets minder dan een derde uit beschrijving van het geldend recht. In de tweede alinea van de noot geeft de annotator aan het eens te zijn met de uitspraak en geeft hij een verklaring waarom de Hoge Raad zo heeft geoordeeld. Hij vergelijkt de onderhavige zaak nog met een eerdere uitspraak, maar concludeert dat de uitspraken op andere kwesties betrekking hebben. De annotator sluit af met het zoeken van een verklaring voor de gewezen beslissing, die eruit bestaat dat de Hoge Raad vermoedelijk een signaal heeft willen afgeven aan het Openbaar Ministerie.

Annotatie nr. 11, arrest Hoge Raad (1 uit 2 goed/uitstekend)

Het arrest verdient volgens de annotator een korte signalering. De noot is dan ook kort (twee derde pagina). Zij bestaat voor twee derde deel uit een samenvatting van de uitspraak en voor een kwart uit een beschrijving van het geldende recht aan de hand van andere, vergelijkbare arresten. In de afsluitende regels geeft de annotator aan het met de uitspraak eens te zijn.

Annotatie nr. 12, arrest Hoge Raad (1 uit 2 goed/uitstekend)

De noot is ruim drie pagina's lang. De annotator begint met het schetsen van problemen die zich geregeld voordoen ten aanzien van de kwestie die in de uitspraak aan de orde is. Na een vergelijking met eerdere jurisprudentie volgt een halve pagina waarin de uitspraak wordt samengevat. Op de daarop volgende 2,5 pagina wordt steeds afwisselend een klein deel van de uitspraak samengevat en door de annotator aangegeven of hij het hier wel of niet mee eens is. De annotator sluit zich overwegend aan bij de uitspraak en uit weinig kritiek. Zeer beperkt wordt ook het geldende recht besproken. Eenmaal wordt vergeleken met Straatsburgse jurisprudentie. Afgesloten wordt met een verwachting voor de toekomst en met het aanstippen van het belang van de uitspraak.

Annotatie nr. 13, arrest Hoge Raad (1 uit 2 goed/uitstekend)

Deze noot van ruim twee pagina's bestaat voor driekwart pagina uit een beschrijving van het geldend recht. Iets minder dan een halve pagina bestaat uit een samenvatting van de uitspraak. Het arrest illustreert volgens de annotator een moeilijkheid. Verder is een halve pagina besteed aan vergelijking met (buitenlandse) jurisprudentie. Meer concreet schetst de annotator een aantal buitenlandse uitspraken waarin een vergelijkbare kwestie aan de orde is. Verder dan het signaleren, noemen en beschrijven van deze gevallen gaat de annotatie echter niet. Een aantal keer geeft de annotator onderbouwd aan het niet geheel eens te zijn met de uitspraak. Op andere plaatsen in de noot geeft hij ook aan waar hij het wel mee eens is.

Annotatie nr. 14, arrest Hoge Raad (0 uit 2 goed/uitstekend)

De annotatie bevat hoofdzakelijk (driekwart van de noot) een vrij uitvoerige schets van achtergronden en een beschrijving van het geval. Daarnaast constateert de annotator dat de uitspraak aansluit bij de heersende leer, maar dat van rechtszekerheid geen sprake is. Afgesloten wordt met het opwerpen van vragen die in de ogen van de annotator belangrijk zijn en door de verwijzingsrechter dienen te worden beantwoord. Deze vragen worden niet beantwoord en ook niet verder uitgewerkt.

Annotatie nr. 15, arrest Hoge Raad (0 uit 2 goed/uitstekend)

Deze noot van 2,5 pagina bestaat nagenoeg helemaal uit een combinatie van een samenvatting van de uitspraak en een beschrijving van het geldende recht. Aan het slot worden enkele vragen opgeworpen (één achtste pagina), maar van beantwoording of een aanzet daartoe is geen sprake.

Annotatie nr. 16, arrest Hoge Raad (0 uit 2 goed/uitstekend)

Deze noot is relatief kort qua omvang, ongeveer een halve pagina. De noot opent met het belang van de uitspraak. Daarna volgt een vergelijking met eerdere jurisprudentie. Nog één zin wordt gewijd aan een overweging van de Hoge Raad in de betreffende zaak. Voor het overige is de noot een beschrijving van geldend recht, onder verwijzing naar literatuur.

Systematische analyse van annotaties (per rechtsgebied)

1 Analyse van bestuursrechtelijke annotaties

De aan- en afwezigheid van kenmerken zijn in een matrix geplaatst om verschillen tussen de annotaties te visualiseren. De matrix hierna laat zien hoe de annotaties zijn beoordeeld en welke kenmerken zij hebben (**Error! Reference source not found.**). Door de annotaties nader met elkaar te vergelijken kan worden gezien of er regelmatigheid te ontdekken zijn die het verschil in beoordelingen kunnen verklaren.

Hoewel een aantal patronen lijkt te bestaan op basis waarvan verschillen tussen goede/uitstekende en matige/slechte annotaties kunnen worden verklaard, zijn er ook annotaties die grotendeels dezelfde kenmerken hebben maar toch verschillend worden gewaardeerd. Het laatste is bijvoorbeeld het geval bij annotatie nr. 16 en annotatie nr. 10. Een nadere bestudering van de annotaties is dan ook op zijn plaats.

Drie annotaties zijn niet in de analyse betrokken. Reden hiervoor is dat één annotatie geen commentaar bij een uitspraak betreft, maar een commentaar is dat bij een rapport is verschenen. De andere twee annotaties zijn geschreven door iemand die tevens de beoordelaar is. Om eventuele vooringenomenheid te voorkomen, zijn deze annotaties buiten beschouwing gelaten.

Tabel 17 – Matrix bestuursrechtelijke annotaties

Score (... van de ... beoordeelaars merkt de annotatie als goed/uitstekend aan)														
22.	Kritiek op voorstellen vanuit de politiek													
21.	Vervachting ontwikkeling	0.5												
20.	Bespreking processtrategie													
19.	Onduidelijkheid / motiveringsgebrek													
18.	Annotator eens / oneens met uitspraak													
17.	Annotator zoekt verklaring													
16.	argumenten voor en tegen de beslissing													
15.	Evaluatie adhv andere uitkomsten													
14.	Evalueren beslissing adhv de praktijk													
13.	Toetsen aan 'hoger' recht													
12.	Toetsen aan doel van de regeling													
11.	Toetsen aansluiting eerdere uitspraken													
10.	Opvating van andere auteur(s)													
9.	Implicaties													
8.	Criterium ontwikkelen en toetsen													
7.	Nagaan welke leer is toegepast													
6.	Opwerpen vragen													
5.	Schetsen van kwesties													
4.	Belang uitspraak													
3.	Beschrijving geldend recht													
2.	Beschrijving feiten/uitspraak													
1.	Hoogste rechter													
ID														
2														
3														
4														
5														
6														
8														
9														
10														
11														
12														
13														

15	1	1	1		1	0 van de 2 (0.0)
16	1	1	1			0 van de 2 (0.0)
17	1	1	1			0 van de 2 (0.0)
18	1	1	1	0.5	1	1 van de 2 (0.5)

Noot: Een score van '1' staat voor aanwezigheid van het betreffende kenmerk. Een score van '0.5' betekent dat het kenmerk is aangetroffen, maar dat de passage(s) van zeer geringe lengte zijn (één of enkele regels) in vergelijking met andere annotaties waarin het kenmerk is geobserveerd.

Toelichting bij categorieën (nummers corresponderen met nummers in tabel op kolomniveau)

1. Uitspraak gewezen door hoogste rechter
2. Beschrijving/samenvatting casus en/of uitspraak
3. Beschrijving geldend recht / verwijzen naar literatuur of jurisprudentie daarover
4. Wijzen op belang uitspraak voor rechtspraktijk of –theorie
5. Schetsen van kwesties die zich geregeld (in de praktijk) voordoen
6. Opwerpen vragen (zonder deze te beantwoorden)
7. Nagaan welke leer is toegepast (welke leer geldend recht is)
8. Criterium/criteria ontwikkelen en toetsen of de uitspraak aan dit criterium voldoet
9. Implicaties van de uitspraak bespreken (bv. voor toepassing/interpretatie geldend recht)
10. Opvatting van andere auteur(s) toetsen / ter discussie stellen
11. Toetsen of uitspraak aansluit bij eerdere uitspraken
12. Toetsen of beslissing van de rechter beantwoordt aan doel van de regeling
13. Toetsen van de uitspraak aan 'hogere' recht (beginselen / verdragsbepalingen / EHRM-rechtspraak)
14. Evalueren beslissing aan de hand van wat in de praktijk (van procesvoering) gebeurt / gangbaar is / gebruikelijk is
15. Evaluatie van de uitspraak door andere denkbare uitkomsten af te zetten tegen de gekozen uitspraak
16. Systematisch de argumenten voor en tegen de beslissing nalopen
17. Annotator geeft (onderbouwd) aan of hij/zij het eens dan wel oneens is met de uitspraak
18. Annotator zoekt naar verklaring voor uitkomst of gewezen rechtsregel
19. Onduidelijke passages constateren c.q. ophelderen / motiveringsgebrek opmerken
20. Aangeven of een procespartij beter een andere keuze had kunnen maken in het proces (bv. andere rechtsgrond invoeren) c.q. wat (andere) mogelijkheden zijn om rechtsposities te versterken / wat er zou zijn gebeurd als een procespartij een andere keuze zou hebben gemaakt
21. Verwachting voor ontwikkeling van leerstuk (volgt niet direct uit uitspraak, anders gecodeerd als 'schetsen implicaties')
22. Kritiek op voorstellen vanuit de politiek

Overzichten/samenvattingen/beschrijvingen

Zoals uit de matrix blijkt, zijn annotaties nr. 16 en nr. 17 als matig/slecht beoordeeld en bestaan zij uitsluitend uit het beschrijven dan wel samenvatten van de casus, de uitspraak, het geldende recht dat daarop van toepassing is en/of het verwijzen naar literatuur of jurisprudentie over dat geldende recht. Annotatie nr. 10 bevat echter dezelfde kenmerken, maar deze annotatie is door beide beoordelaars als goed/uitstekend aangemerkt. De reden hiervoor moet vermoedelijk worden gevonden in het handzame overzicht dat annotatie nr. 10 biedt. Eén beoordelaar voerde dit ook expliciet als reden aan voor de positieve beoordeling. Hij spreekt in dit verband van een ‘mooie update’.

Diezelfde reden wordt door deze beoordelaar aangevoerd om annotaties nr. 5 en nr. 8 als goed/uitstekend aan te merken, terwijl deze annotaties door de andere beoordelaar als matig/slecht worden aangemerkt. Het valt bij deze twee annotaties niet uit te sluiten dat de positieve waardering van die ene beoordelaar (mede) te maken heeft met de aanwezigheid van andere kenmerken, zoals het toetsen of de beslissing van de rechter beantwoordt aan het doel van de regeling (nr. 5) of het evalueren van de beslissing aan de hand van wat in de praktijk gangbaar is (nr. 8).

Ook annotatie nr. 11 bestaat uitsluitend uit een weergave van de feiten, het geldende recht en/of het geldende recht dat op de problematiek betrekking heeft. Net als bij annotaties nr. 5 en nr. 8 merkt één van de twee beoordelaars deze annotatie aan als goed/uitstekend. Het kan aan de hand van de ontvangen toelichting niet worden achterhaald hoe deze beoordelaar is gekomen tot deze waardering.

Het blijkt niet mogelijk om indicatoren aan te wijzen die verklaren onder welke omstandigheden een annotatie die geheel of hoofdzakelijk bestaat uit een weergave van de casus, de uitkomst en/of het geldende recht dat op dergelijke gevallen van toepassing is, als goed/uitstekend wordt aangemerkt, en wanneer als matig/slecht. Positieve en minder positieve beoordelingen ten aanzien van deze kenmerken lijken vooral samen te hangen met de opvatting van de beoordelaar.

Samengevat lijkt het weergeven van de feiten, de uitspraak en het geldende recht dat op het geval van toepassing is soms wel en dan weer niet te kunnen worden gerelateerd aan de kwaliteit van een annotatie. Over het algemeen geldt dat annotaties waarin dit uitsluitend of in overwegende mate aan de orde is, niet snel als goed/uitstekend worden aangemerkt, tenzij er bij de lezer (in dit geval de beoordelaars) een zekere behoefte bestaat aan een overzicht. Het laatste is echter subjectief en daarom moeilijk vast te stellen.

Beschrijving verschillende opvattingen in literatuur

Het beschrijven van verschillende opvattingen is aan de orde in annotatie nr. 5 en annotatie nr. 13. In annotatie nr. 5, die door één beoordelaar als goed/uitstekend is aangemerkt en door de andere als matig/slecht, maakt de beschrijving onderdeel uit van het toetsen of de beslissing beantwoordt aan het doel van de regeling. Voor annotatie nr. 13 kan, mede gezien de geringe omvang van de passages en de aanwezigheid van andere kenmerken, niet worden vastgesteld dat er een relatie bestaat tussen het beschrijven van verschillende opvattingen en de kwaliteit van een annotatie. Een positieve relatie is echter zeer onwaarschijnlijk gezien de negatieve beoordeling van deze noot; beide beoordelaars kwalificeerden deze noot als matig/slecht. Op grond hiervan stel ik vast dat het beschrijven van opvattingen uit de literatuur op zichzelf waarschijnlijk geen indicator kan zijn voor de kwaliteit van een (bestuursrechtelijke) annotatie.

Toetsen aan 'hoger' recht

Annotaties waarin is getoetst aan 'hoger' recht, waaronder beginselen of verdragsbepalingen, lijken in positieve zin bij te dragen aan de kwaliteit van (bestuursrechtelijke) annotaties. Zo geldt voor de drie annotaties waarin het toetsen aan 'hoger' recht aan de orde is (nr. 3, nr. 6 en nr. 9), dat zij als goed/uitstekend zijn aangemerkt door beide beoordelaars. Van de overige twee annotaties die als goed/uitstekend zijn aangemerkt, maar waarin niet wordt getoetst aan 'hoger' recht (nr. 2 en nr. 10), kan om een andere reden worden verklaard dat zij deze score hebben gekregen.¹⁶⁴

Dat het toetsen aan 'hoger' recht vermoedelijk een positief effect heeft op de kwaliteit van (bestuursrechtelijke) annotaties blijkt het meest uit annotatie nr. 6. Deze annotatie heeft, behalve het toetsen aan 'hoger' recht en het beschrijven van de feiten, uitkomst en het geldende recht, geen ander kenmerk op basis waarvan de kwaliteit van de annotatie kan worden verklaard. Er zijn geen indicaties om te veronderstellen dat de kwaliteit (vooral) moet worden verklaard door de weergave van de feiten, de uitspraak en/of het geldende recht. Bovendien wijzen beide beoordelaars in hun toelichting bij de gegeven scores op het belang van de analyse en niet (zozeer) op de adequate weergave. Dit doet vermoeden dat de positieve waarderings (vooral) moeten worden verklaard door het toetsen aan 'hoger' recht.

In annotatie nr. 6 gaat de annotator na of de rechter in de besproken uitspraak marginaal had mogen toetsen. De annotator concludeert van niet, omdat marginale toetsing niet aan de orde zou zijn. De rechter had in dit geval marginaal getoetst, met als gevolg dat klaagster een ruime periode na haar subsidieaanvraag nog geen uitspraak had gekregen ten aanzien van het materiële geschil.

¹⁶⁴ Zie hiervoor voor nr. 10 (handzaam overzicht) en hierna voor nr. 2 (schetsen implicaties).

De annotator laat, kort samengevat, zien dat de feiten in dit type zaken relatief eenvoudig zijn vast te stellen, alsook de consequenties daarvan voor het verkrijgen en het vaststellen van de hoogte van de subsidie. Een ‘volle’ toets zou daarmee niet hebben geleid tot een schending van het beginsel van rechterlijke terughoudendheid. Daarom is een marginale toets in dit type gevallen niet nodig, aldus de annotator. Deze redenering beslaat ongeveer de helft van de annotatie.

Ook in annotatie nr. 3 is het toetsen aan ‘hoger’ recht aan de orde. Hier toetst de annotator de uitspraak aan een bepaald beginsel. Hij doet dit mede aan de hand van voorbeelden uit het strafrecht en het privaatrecht. Meer specifiek gaat de annotator na hoe in deze rechtsgebieden om wordt gegaan met botsingen tussen het bedoelde beginsel en andere uitgangspunten. Op basis van deze vergelijking concludeert de annotator dat de oplossing die de rechter in de besproken zaak heeft gekozen, in de toekomst geen navolging verdient.

Het toetsen van de uitspraak aan ‘hoger’ recht beslaat een groot deel van annotatie nr. 3 (twee van de vijf pagina’s). Verder zijn ruim twee pagina’s van deze annotatie gewijd aan een weergave van de feiten, de uitspraak en het geldende recht dat van toepassing is op het besproken geval. Voor het overige bestaat de annotatie uit het schetsen van implicaties, het vergelijken met eerdere jurisprudentie, het opmerken van een motiveringsgebrek en uit het zoeken naar een verklaring voor de gewezen beslissing. Gezien de centrale plaats van het toetsen aan ‘hoger’ recht is het waarschijnlijk dat dit kenmerk een positief effect heeft op de kwaliteit van deze annotatie, hoewel niet uit te sluiten valt dat de andere van de genoemde kenmerken daar (mede) op van invloed zijn geweest.

Annotatie nr. 9 gaat na of een bepaald criterium, dat in de uitspraak is aangescherpt, in overeenstemming is met een verdragsrechtelijke bepaling. De annotator concludeert al vrij snel dat de uitspraak de verdragsrechtelijke toets doorstaat. De passage waarin aan de verdragsrechtelijke bepaling wordt getoetst, is dan ook beperkt qua omvang (een alinea op de hele annotatie, die vier pagina’s telt). Het is daarom onwaarschijnlijk dat alleen dit kenmerk de positieve waardering van annotatie nr. 9 kan verklaren. Vermoedelijk is (ook of vooral) de aanwezigheid van andere kenmerken van belang, zoals het schetsen van de implicaties, dat bovendien uitgebreider aan de orde komt (ongeveer één pagina, zie ook hierna).

Het effect van het toetsen aan ‘hoger’ recht op de kwaliteit van de annotatie is het duidelijkst bij annotatie nr. 6, omdat dit kenmerk het kenmerk is waarmee de annotatie zich onderscheidt van annotaties die als kwalitatief minder goed zijn aangemerkt. Ook uit annotatie nr. 3 valt af te leiden dat het toetsen aan ‘hoger’ recht een indicatie is voor een kwalitatief goede of uitstekende noot. Het effect is minder duidelijk bij annotatie nr. 9, waar het kenmerk een relatief klein deel van de annotatie beslaat.

Het voorgaande levert aanwijzingen op dat het toetsen aan ‘hoger’ recht een positieve invloed heeft op de kwaliteit van annotaties, mits de passages waarin wordt getoetst een substantiële omvang hebben.

Implicaties uitspraak

Behalve het toetsen aan ‘hoger’ recht, lijkt ook het bespreken van implicaties van de uitspraak een indicator te zijn voor de kwaliteit van annotaties. Dit kenmerk komt uitsluitend terug bij annotaties die als goed/uitstekend worden aangemerkt, meer specifiek bij nr. 2, nr. 3 en nr. 9.

Annotatie nr. 2 kwam hiervoor al summier aan de orde, echter toen in relatie tot het toetsen aan ‘hoger’ recht. Een belangrijk deel van de annotatie (een halve pagina op drie pagina’s annotatie) behandelt de consequenties van de uitspraak voor de interpretatie en toepassing van het recht. De vraag die de annotator in dit verband opwerpt, is of de uitspraak meebrengt dat een Nederlandse bestuursrechter ambtshalve dient te toetsen aan een tweetal artikelen uit het EG-verdrag. Dat is niet het geval, aldus de annotator, die vaststelt dat het ambtshalve aanvullen van gronden, waar in het besproken geval sprake van was, nog geen ambtshalve toetsing impliceert en dat zelfs als dit het geval zou zijn, nog niet is gezegd hoe ver de ambtshalve toepassing zou moeten gaan.

Van een andere orde is annotatie nr. 3. Daar worden de implicaties van de uitspraak in een enkele alinea geschetst. Gezien de summiere uitwerking valt niet te verwachten dat deze passage uit annotatie nr. 3 in belangrijke mate heeft bijgedragen aan de (positieve) waarderingen door de beoordelaars. Die lijken, zoals hiervoor besproken, eerder verband te houden met het toetsen aan ‘hoger’ recht.

Annotatie nr. 9 gaat uitgebreider dan nr. 3 in op de implicaties van de uitspraak voor de interpretatie en toepassing van het geldende recht. Zo constateert de annotator dat de Afdeling een strikter criterium hanteert dan voorheen bij de beoordeling of een organisatie als belanghebbende is aan te merken, wat voor organisaties tot gevolg heeft dat zij hun doelstellingen tegen het licht dienen te houden. Verder zoekt de annotator naar een afbakening van het criterium dat door de Afdeling wordt gehanteerd: wanneer is een organisatie belanghebbende? In dit verband verwijst de annotator naar andere uitspraken, onder meer naar een uitspraak van de Hoge Raad. Van een heldere lijn kan volgens de annotator echter nog niet worden gesproken.

Met name uit annotatie nr. 2 en annotatie nr. 9 volgt dat er een positieve relatie bestaat tussen het bespreken van implicaties van de uitspraak en de kwaliteit van annotaties. Verder dienen de passages waarin implicaties worden geschetst, van enige omvang te zijn (meer dan een enkele alinea).

Zoeken naar verklaring voor de gewezen uitspraak

In twee van de onderzochte (bestuursrechtelijke) annotaties zoekt de annotator naar een verklaring voor de gewezen uitspraak, te weten in annotaties nr. 3 en nr. 13. De eerste annotatie (nr. 3) is door beide beoordelaars als goed/uitstekend aangemerkt, de tweede (nr. 13) door beide beoordelaars als matig/slecht. In beide annotaties beslaan de passages waarin naar een verklaring wordt gezocht een klein onderdeel van de annotatie. Beide annotaties hebben tevens gemeenschappelijk dat zij de nodige slagen om de arm houden ten aanzien van de verklaring ('wij kunnen ons voorstellen', 'ik sluit niet uit'). Op grond van deze vergelijking is er onvoldoende reden om aan te nemen dat het zoeken naar een verklaring voor de gewezen beslissing een indicatie is voor een kwalitatief goede of uitstekende annotatie.

Gemotiveerd aangeven of de uitspraak overtuigt

Het geven van voor- en tegenargumenten ten aanzien van de uitspraak in het individuele geval ('de rechter had hier anders moeten beslissen'), lijkt niet van invloed te zijn op de kwaliteit van (bestuursrechtelijke) annotaties. Van de drie annotaties waarin het kenmerk aan de orde was, is er één door beide beoordelaars als matig/slecht (nr. 15) beoordeeld en zijn de overige twee annotaties door één van de twee beoordelaars als goed/uitstekend aangemerkt (nr. 4 en nr. 18). Laatstgenoemde twee annotaties hebben gemeen dat een vergelijking plaatsvindt met eerder gedane uitspraken en dat deze uitspraken in het betoog zijn betrokken, maar er is onvoldoende grond om hier conclusies aan te verbinden. Op grond van de onderzochte (bestuursrechtelijke) annotaties zijn er dan ook geen aanwijzingen om te veronderstellen dat het gemotiveerd aangeven of de uitspraak overtuigt een effect heeft op de kwaliteit van (bestuursrechtelijke) annotaties.

Vergelijken met eerdere uitspraken

Het vergelijken van de geannoteerde uitspraak met andere, eerder gedane uitspraken lijkt geen indicator te zijn voor de kwaliteit van (bestuursrechtelijke) annotaties. Zowel in de goede als minder goede annotaties komt het vergelijken met eerdere uitspraken terug.

De indruk die ontstaat bij een nadere bestudering van de annotaties is dat de relatie tussen het vergelijken met eerdere jurisprudentie en de kwaliteit van een annotatie afhangt van de wijze waarop wordt vergeleken of met welk doel de vergelijking plaatsvindt. De gegevens in de matrix suggereren dat het vergelijken met eerdere uitspraken de kwaliteit van annotaties vergroot indien de vergelijking gepaard gaat met het toetsen aan 'hogere' recht en/of het bespreken van de implicaties van de uitspraak voor het geldende recht en de toepassing ervan. Indien voor de geannoteerde uitspraak bijvoorbeeld wordt bekeken of

deze aansluit bij rechtspraak van het *Europees Hof voor de Rechten van de Mens*, vergroot dit de kans op een kwalitatief goede annotatie. Het hangt er uiteraard wel van af *hoe* wordt vergeleken, zo bleek eerder uit de vergelijking tussen annotatie nr. 2 en nr. 13. Indien de vergelijking bijvoorbeeld blijft steken bij de constatering dat de uitspraken wel of niet aansluiten ('de rechter bevestigt eerdere rechtspraak'), werkt dit vermoedelijk niet kwaliteitsverhogend. De annotator zal na de vergelijking de implicaties van de uitspraak moeten laten zien of aan moeten geven hoe een bepaald beginsel wordt uitgewerkt en waarom de uitspraak in kwestie daar al dan niet bij aansluit. In die gevallen gaat het vergelijken met eerdere uitspraken echter op in bijvoorbeeld het toetsen aan 'hoger' recht of in het schetsen van implicaties.

Wijzen op het belang van de uitspraak

In een aantal annotaties wordt gewezen op het belang van de uitspraak voor het recht of de rechtspraak. Dit geldt voor annotatie nr. 2, nr. 9 en nr. 11, waarvan de eerste twee noten door beide beoordelaars als goed/uitstekend zijn aangemerkt en de laatstgenoemde annotatie door één beoordelaar als goed/uitstekend. Het ligt niet voor de hand dat het wijzen op het belang van de uitspraak op zichzelf leidt tot een hogere kwaliteit van de annotatie. Het is waarschijnlijker dat de kwaliteit veel meer wordt beïnvloed door de kenmerken die ik eerder omschreef. Steun hiervoor kan worden gevonden in de beperkte omvang van de passages waarin op het belang van de uitspraak wordt gewezen in relatie tot de omvang van de passages waarin andere kenmerken aan de orde zijn.

Toetsen of beslissing beantwoordt aan doel van de regeling

In twee van de bestudeerde annotaties (nr. 5, nr. 12) gaat de annotator na of de uitkomst beantwoordt aan het doel van een bepaalde regel of regeling. Voor annotatie nr. 12 ligt het voor de hand dat dit kenmerk één beoordelaar ertoe heeft bewogen om de noot als goed/uitstekend aan te merken. De beschrijving van de feiten en de uitspraak is dermate kort dat dit de positieve beoordeling vermoedelijk moeilijk kan verklaren. Verder ontbreken andere kenmerken die het verschil zouden kunnen verklaren. Hetzelfde geldt voor annotatie nr. 5, hoewel daar het beschrijvende deel groter is. Omdat dezelfde beoordelaar beide noten als goed/uitstekend heeft aangemerkt, en de andere beoordelaar beide noten als matig/slecht heeft gekwalificeerd, is het waarschijnlijk dat het toetsen of een beslissing beantwoordt aan het doel van de regeling voor de ene beoordelaar wel, maar voor de andere beoordelaar niet een criterium is om (bestuursrechtelijke) annotaties te beoordelen.

Onduidelijke passages constateren c.q. ophelderen/motiveringsgebrek opmerken

In twee annotaties (nr. 3, nr. 4) merkt de annotator een motiveringsgebrek op ('de rechter had er wel wat meer woorden aan mogen wijden', 'de rechter motiveert niets'). Hiervan wordt de ene door beide beoordelaars als goed/uitstekend aangemerkt (nr. 3) en de andere door één beoordelaar (nr. 4). Voor beide annotaties geldt echter dat de passages over het motiveringsgebrek kort zijn (circa acht regels), dat andere kenmerken dominanter zijn en dat er geen poging wordt gedaan om het motiveringsgebrek op te helderen. Op grond hiervan is het onwaarschijnlijk dat het opmerken van een motiveringsgebrek een positieve invloed heeft op de kwaliteit van de onderzochte annotaties.

Verwachting ontwikkeling leerstuk

In twee annotaties (nr. 2, nr. 10), die door beide beoordelaars als goed/uitstekend zijn aangemerkt, schetst de annotator een verwachting voor de ontwikkeling van het leerstuk. In annotatie nr. 2 blijft het bij aanstippen. Daar stelt de annotator, onder verwijzing naar beleidsstukken, dat de invloed van Europa ten aanzien van een bepaald leerstuk 'weleens verder zou kunnen toenemen'. Ook omdat andere kenmerken dominanter zijn in deze annotatie dan het schetsen van een toekomstverwachting, is het niet waarschijnlijk dat het uitspreken van deze verwachting een wezenlijke rol heeft gespeeld voor de kwaliteit van deze noot.

In annotatie nr. 10 komt het kenmerk uitgebreider en op een andere wijze aan de orde. Daar schetst de annotator de kritiek die is geuit op hoe de rechter omgaat met een bepaald probleem. Die kritiek lijkt niet gering, aldus de annotator, en gezien de kracht van de argumenten verwacht hij daarom dat het niet lang meer zal duren voordat de rechter anders zal optreden. Omdat annotatie nr. 10 onder te verdelen is in twee kenmerken (weergave van feiten/uitspraak/geldend recht en het uitspreken van een verwachting voor de toekomst), kan in dit geval niet worden vastgesteld dat het uitspreken van een toekomstverwachting verklaart waarom deze uitspraak positief is beoordeeld door beide beoordelaars. Evenwel zouden de positieve beoordelingen kunnen worden toegeschreven aan de adequate weergave van de casus, de uitspraak en het geldende recht dat op het geval van toepassing is. Mocht het uitspreken van een toekomstverwachting een positief effect hebben gehad op de kwaliteit van deze annotatie, dan is de verklaring daarvoor waarschijnlijk gelegen in de omstandigheid dat de verwachting wordt onderbouwd aan de hand van kritiek die eerder is geuit (of kan worden geuit) op de huidige praktijk of regel.

Overige kenmerken

Een aantal annotaties bevat een kenmerk dat alleen in die annotatie terugkomt. In annotatie nr. 9 gaat het om het uiten van kritiek op voorstellen van Kamerleden ten aanzien van een bepaald juridisch leerstuk. Dit kenmerk betreft één ali-

nea van de gehele annotatie, waarin bovendien met name het schetsen van implicaties dominant is. Vermoedelijk verklaart vooral dit laatste kenmerk de positieve scores die beide beoordelaars aan de annotatie verbonden. Niet valt uit te sluiten dat het uiten van kritiek op voorstellen uit de politiek heeft bijgedragen aan de positieve scores, maar gezien de omvang van de passage is de invloed ervan, als die er al is, vermoedelijk beperkt.

In annotatie nr. 8 evalueert de annotator de beslissing aan de hand van wat in de praktijk (van procesvoering) gangbaar en gebruikelijk is, nadat hij de uitkomst en het geldende recht dat op het geval van toepassing is heeft beschreven. De annotator stelt en illustreert in dit verband dat de hoogste rechter in dit geval eist van de voorzieningenrechter om ergens aandacht aan te besteden als rechters in bodemzaken dat ook niet doen. Vanwege de afwezigheid van andere kenmerken is de aanwezigheid van dit kenmerk voor één van de twee beoordelaars vermoedelijk de reden geweest om de noot als goed/uitstekend te kwalificeren. De andere beoordelaar heeft de noot als matig/slecht gekwalificeerd, zodat, mede gezien de omstandigheid dat het kenmerk niet in andere annotaties terugkomt,¹⁶⁵ er onvoldoende reden is om dit kenmerk als kwaliteitsindicator te beschouwen.

Conclusie

Annotaties die voor het overgrote deel of uitsluitend bestaan uit een weergave van de feiten, het oordeel van de rechter en/of het geldende recht dat op de problematiek van toepassing is, maakt een annotatie in een enkel geval, maar over het algemeen niet, tot een kwalitatief goede of uitstekende annotatie. Indien een dergelijke annotatie wel als goed/uitstekend wordt aangemerkt, lijkt dit te maken te hebben met het bruikbare overzicht dat het geeft. Of daar sprake van is, is moeilijk aan te geven. Het louter weergeven van feiten, uitkomsten en het geldende recht is dan ook een belangrijke indicatie voor een matige of zelfs slechte annotatie.

Indicatoren die positief zijn geassocieerd met de kwaliteit van (bestuursrechtelijke) annotaties, zijn het toetsen aan ‘hoger’ recht, het schetsen van implicaties voor de interpretatie van het geldende recht en/of de toepassing ervan. Zonder uitzondering bestaat de kern van de onderzochte annotaties, naast een beschrijving van de casus, uitspraak en/of het geldende recht, uit een van deze kenmerken of een combinatie ervan. De passages waarin de kenmerken aan worden getroffen, beslaan samen of afzonderlijk een substantieel deel van de annotatie.

Van het schetsen van een verwachting voor de toekomstige ontwikkeling van het (bestuurs)recht is onduidelijk of dit kenmerk een positief effect heeft op

¹⁶⁵ Ook niet in privaatrechtelijke of strafrechtelijke annotaties.

de kwaliteit van annotaties. Als dat het geval is, dient de verwachting die de annotator schetst grondslag of rechtvaardiging te vinden in de literatuur of andere betrouwbare bronnen waarin (bijvoorbeeld) kritiek op de huidige regel of toepassing is geuit.

De analyse van bestuursrechtelijke annotaties levert verder op dat het beschrijven van opvattingen uit de literatuur en het wijzen op het belang van de uitspraak op zichzelf niet leidt tot een hogere kwaliteit van de annotatie. Voorts is er op basis van de analyse geen reden te veronderstellen dat er een relatie is tussen de kwaliteit van een (bestuursrechtelijke) annotatie en het zoeken naar een verklaring voor de gewezen beslissing, het gemotiveerd aangeven of de uitspraak overtuigt, het vergelijken met eerdere uitspraken, het toetsen of de beslissing beantwoordt aan het doel van de regeling, het opmerken van een motiveringsgebrek, kritiek op voorstellen uit de politiek en het evalueren van een uitspraak aan de hand van wat gangbaar of gebruikelijk is in de praktijk.

2 Analyse van privaatrechtelijke annotaties

Net als bij de bestuursrechtelijke annotaties, suggereert het overzicht dat bepaalde kenmerken kunnen worden gerelateerd aan de scores van privaatrechtelijke annotaties zoals gegeven door de (drie) beoordelaars (**Error! Reference source not found.**). Een bespreking van de verschillende kenmerken volgt hierna. Eén annotatie is niet in de analyse betrokken. Twee van de drie beoordelaars plaatsen bij deze noot de kanttekening dat zij geen annotatie was, maar een artikel.

Overzichten/samenvattingen/beschrijvingen

Uit het overzicht volgt dat annotatie nr. 13, die door alle drie de beoordelaars als matig/slecht is gekwalificeerd, niet verder gaat dan het weergeven van de feiten, de overwegingen, de uitkomst en het recht dat op het besproken geval van toepassing is. Een vergelijkbare conclusie kan worden getrokken voor annotatie nr. 1. Weliswaar bevat deze annotatie ook vergelijkingen met andere uitspraken en wordt in de annotatie op het belang van de uitspraak gewezen, maar deze passages beslaan slechts enkele regels op de totale annotatie, zodat het effect, als het er al is, vermoedelijk beperkt is. De analyse heeft voorts geen aanleiding gegeven te veronderstellen dat de annotaties nr. 1 en nr. 13 qua beschrijving, weergave of leesbaarheid wezenlijk afwijken van andere annotaties. De opbouw of inhoud van deze annotaties, alsook het commentaar van de beoordelaars, gaf overigens geen aanleiding om te veronderstellen dat annotaties nr. 1 en nr. 13 qua beschrijving, weergave of leesbaarheid afwijken van andere annotaties. Op basis hiervan kan worden aangenomen dat het uitsluitend of overwegend weergeven van feiten, de uitkomst en/of het geldende recht een indicatie is voor een matige of slechte privaatrechtelijke annotatie.

Tabel 18 – Matrix privaatrechtelijke annotaties

Score (... van de ... beoordeelaars merkt de annotatie als goed/uitstekend aan)							
22. Kritiek op voorstellen vanuit de politiek	0	van	de	3			
21. Verwachting ontwikkeling	3	van	de	3			
20. Bespreking processtrategie	0	van	de	3			
19. Onduidelijkheid / motiveringsgebrek	3	van	de	3			
18. Annotator eens / oneens met uitspraak	1	van	de	3			
17. Annotator zoekt verklaring	0.5						
16. Argumenten voor en tegen de beslissing							
15. Evaluatie adhv andere uitkomsten	1						
14. Evalueren beslissing adhv de praktijk							
13. Toetsen aan 'hoger' recht							
12. Toetsen aan doel van de regeling							
11. Toetsen aansluiting eerdere uitspraken	0.5						
10. Opvatting van andere auteur(s)							
9. Implicaties	1						
8. Criterium ontwikkelen en toetsen							
7. Nagaan welke leer is toegepast							
6. Opwerpen vragen							
5. Schetsen van kwesties							
4. Belang uitspraak	1						
3. Beschrijving geldend recht	1						
2. Beschrijving feiten/uitspraak	1						
1. Hoogste rechter	1						
ID	1						

7	1	1	1	1		0.5	0.5									1 van de 3 (0.333)
8	1	1	1			1	0.5	0.5			0.5	0.5				2 van de 3 (0.667)
9	1	1	1			0.5	0.5									0 van de 3 (0.0)
10	1	1	1			1	0.5									2 van de 3 (0.667)
11	1	1	1			1		0.5	0.5							1 van de 3 (0.333)
12	1	1	1			1	0.5	0.5		1		0.5				2 van de 3 (0.667)
13			1	1												0 van de 3 (0.0)
14	1	1	1					0.5				1				1 van de 3 (0.333)
15			1	1				0.5				0.5				0 van de 2 (0.0)

Noot: Een score van '1' staat voor aanwezigheid van het betreffende kenmerk. Een score van '0.5' betekent dat het kenmerk is aangetroffen, maar dat de passage(s) van zeer geringe lengte zijn (één of enkele regels) in vergelijking met andere annotaties waarin het kenmerk is geobserveerd. *Score gebaseerd op twijfels onder de beoordelaars of het schetsen van implicaties in dit geval reëel was.

Toelichting bij categorieën (nummers corresponderen met nummers in tabel op kolomniveau)

1. Uitspraak gewezen door hoogste rechter
2. Beschrijving/samenvatting casus en/of uitspraak
3. Beschrijving geldend recht / verwijzen naar literatuur of jurisprudentie daarover
4. Wijzen op belang uitspraak voor rechtspraktijk of –theorie
5. Schetsen van kwesties die zich geregeld (in de praktijk) voordoen
6. Opwerpen vragen (zonder deze te beantwoorden)
7. Nagaan welke leer is toegepast (welke leer geldend recht is)
8. Criterium/criteria ontwikkelen en toetsen of de uitspraak aan dit criterium voldoet
9. Implicaties van de uitspraak bespreken (bv. voor toepassing/interpretatie geldend recht)
10. Opvatting van andere auteur(s) toetsen / ter discussie stellen
11. Toetsen of uitspraak aansluit bij eerdere uitspraken
12. Toetsen of beslissing van de rechter beantwoordt aan doel van de regeling
13. Toetsen van de uitspraak aan 'hogere' recht (beginselen / verdragsbepalingen / EHRM-rechtspraak)
14. Evalueren beslissing aan de hand van wat in de praktijk (van procesvoering) gebeurt / gangbaar is / gebruikelijk is
15. Evaluatie van de uitspraak door andere denkbare uitkomsten af te zetten tegen de gekozen uitspraak
16. Systematisch de argumenten voor en tegen de beslissing nalopen
17. Annotator geeft (onderbouwd) aan of hij/zij het eens dan wel oneens is met de uitspraak
18. Annotator zoekt naar verklaring voor uitkomst of gewezen rechtsregel
19. Onduidelijke passages constateren c.q. ophelderen / motiveringsgebrek opmerken
20. Aangeven of een procespartij beter een andere keuze had kunnen maken in het proces (bv. andere rechtsgrond invoeren) c.q. wat (andere) mogelijkheden zijn om rechtsposities te versterken / wat er zou zijn gebeurd als een procespartij een andere keuze zou hebben gemaakt
21. Verwachting voor ontwikkeling van leerstuk (volgt niet direct uit uitspraak, anders gecodeerd als 'schetsen implicaties')
22. Kritiek op voorstellen vanuit de politiek

Vergelijken met eerdere uitspraken

Het vergelijken met andere, eerdere uitspraken blijkt bij privaatrechtelijke annotaties geen onderscheidend criterium te zijn aan de hand waarvan de kwaliteit kan worden bepaald. Bij 12 van de 15 geanalyseerde privaatrechtelijke annotaties was sprake van een vergelijking. Dit aspect komt terug bij zowel annotaties die (overwegend) als goed/uitstekend werden aangemerkt (zie bijvoorbeeld nr. 2, nr. 8 en nr. 12) als bij annotaties die overwegend als matig/slecht werden aangemerkt (nr. 1, nr. 3, nr. 9, nr. 15). Inhoudelijk valt dit mogelijk te verklaren door de relatieve rol die deze vergelijkingen spelen in annotaties. Zonder uitzondering gaat het bij de vergelijkingen om de constatering, vaak in één of enkele zinnen, dat een uitspraak al dan niet aansluit bij eerdere jurisprudentie, soms aangevuld met een korte beschrijving van de uitspra(a)k(en) waarmee wordt vergeleken.

Implicaties van de uitspraak

In 11 van de 15 geanalyseerde annotaties is het beschrijven van de implicaties aan de orde. Dit geldt voor zowel annotaties die (overwegend) als goed/uitstekend zijn gekwalificeerd (nr. 2, nr. 4, nr. 8, nr. 10, nr. 12) als voor annotaties die dat niet zijn (met name nr. 3 en nr. 9, maar ook nr. 5, nr. 6, nr. 7, nr. 11).

Een nadere analyse leverde op dat er verschillen bestaan tussen de annotaties waarin het bespreken van de implicaties aan de orde was. Bij een aantal van de annotaties dat als matig/slecht is aangemerkt (nr. 3, nr. 6, nr. 7, nr. 9) valt op dat het bespreken van de implicaties slechts enkele regels in beslag neemt. Het gaat daarbij om uitspraken als '[...] ligt het voor de hand dat [iemand iets] moet bewijzen. Meer gevolgen zou ik aan de overweging niet willen verbinden', of '[h]et is moeilijk situaties te bedenken waarin het criterium toepassing zou kunnen vinden'. Een ander voorbeeld is de annotatie waarin de annotator beschrijft tot welke uitkomst de toepassing van de rechtsregel op een ander leerstuk leidt, maar ook hier beperkt het voorbeeld zich tot enkele regels. Bovendien merkt de annotator in deze annotatie op dat zijn voorspelling afwijkt van uitspraken in rechtspraak over de problematiek, wat verder niet nader wordt besproken of wordt toegelicht.

De implicaties die in annotatie nr. 11 worden besproken, hebben een meer substantiële omvang (twee alinea's). Hierin beschrijft de annotator dat de rechter, net als in een eerdere uitspraak, de deur heeft dichtgezet voor het vergoeden van een bepaald type schade. Om dit te doorbreken, zo vervolgt de annotator, is het noodzakelijk dat partijen andere omstandigheden of factoren aanvoeren, waarvan de annotator er een aantal noemt.

Deze bespreking van de implicaties van de uitspraak heeft niet tot een positieve waardering door de beoordelaars geleid, zo laat het overzicht zien. Moge-

lijk heeft dit te maken met de omstandigheid dat de consequentie, dat een eiser bijzondere omstandigheden dient aan te voeren om een bepaald type schade vergoed te krijgen, voor de hand ligt. En hoewel de annotator een aantal mogelijke gezichtspunten noemt, werkt hij die niet verder uit. Daarmee zou het schetsen van implicaties in dit geval te 'licht' kunnen zijn geweest. Uit de toelichtingen van de beoordelaars valt daar niets over op te maken.

Een bijzonder geval betreft annotatie nr. 5. Deze annotatie bevat meerdere alinea's waarin de annotator de gevolgen van de uitspraak voor andere leerstukken bespreekt. De invulling die in de uitspraak wordt gegeven aan een bepaald criterium is volgens de annotator tevens relevant voor andere leerstukken omdat daar een vergelijkbaar criterium wordt gebruikt. De annotator geeft hierbij echter zelf aan dat vraagtekens kunnen worden geplaatst bij de vergelijkbaarheid van de rechtsfiguren, en dus ook bij het doortrekken van conclusies van het ene rechtsgebied naar het andere rechtsgebied. Uit de toelichting die de beoordelaars hebben gegeven, blijkt dat deze omstandigheid voor ten minste één van de beoordelaars reden is geweest om de annotatie niet als goed/uitstekend aan te merken. Volgens deze beoordelaar trok de annotator onterecht de conclusies door naar andere leerstukken. Dit verklaart waarschijnlijk waarom deze annotatie slechts door één beoordelaar als goed/uitstekend is aangemerkt.

Het bespreken van de implicaties van de uitspraak komt uitgebreider aan de orde in de twee annotaties die door alle beoordelaars als goed/uitstekend zijn aangemerkt (nr. 2 en nr. 4), in vergelijking met de meeste van de eerder genoemde annotaties (met name nr. 6, nr. 7 en nr. 9). Een belangrijk deel van annotatie nr. 2 gaat over de gevolgen van de uitspraak. De eerste keer dat die worden besproken, is als de annotator de vraag stelt welk moment beslissend is om bepaalde wetenschap bij betrokkenen te beoordelen. Hij gaat op zoek naar dat moment door (fictieve)¹⁶⁶ gevallen te beschrijven en door na te gaan wat redelijk is gezien de uitkomsten in die gevallen. De tweede keer dat implicaties van de uitspraak worden besproken, is als de annotator zich de vraag stelt of het criterium zoals dat door de rechter is gebruikt, leidt tot resultaten die redelijk zijn. Aan de hand van verschillende, zelfbedachte casus oordeelt de annotator dat dit niet het geval is. De derde keer dat de annotator de gevolgen van de uitspraak aan de orde stelt, is wanneer hij constateert dat de Hoge Raad een bepaald criterium blijkt toe te passen, waarna hij aan de hand van voorbeelden (nieuwe situaties) illustreert hoe het criterium dient te worden toegepast.

De passage waarin in annotatie nr. 4 de implicaties van de uitspraak worden geschetst, is minder omvangrijk dan in bijvoorbeeld annotatie nr. 2. Zij beslaat ruim een alinea. In een poging om de relevante omstandigheden te achterhalen waarop de annotator zijn uitkomsten baseert, bespreekt de annotator

¹⁶⁶ Niet ontleend aan jurisprudentie.

wat er zou zijn gebeurd als de feiten anders zouden hebben gelegen. De annotator betoogt dat de uitkomst dan niet anders zou zijn geweest en duidt daarbij de omstandigheid die maakt dat het resultaat hetzelfde blijft. Hiermee distilleert de annotator als het ware de relevante gezichtspunten die gebruikt dienen te worden om gevallen te beslechten zoals in de uitspraak aan de orde.

Ook annotatie nr. 10 bespreekt de implicaties van de uitspraak. Behalve dat de annotator in een alinea de gevolgen voor het waarden van het bewijs bespreekt, gaat hij tevens op zoek naar gezichtspunten aan de hand waarvan situaties vergelijkbaar aan die in de uitspraak moeten worden beoordeeld. Hij komt tot die gezichtspunten door de besproken uitspraak te vergelijken met eerdere uitspraken en de gezichtspunten die in die uitspraken naar voren komen met elkaar te combineren. Deze passages beslaan twee alinea's.

Annotaties nr. 8 en nr. 12 zijn annotaties die overwegend als goed/uitstekend worden aangemerkt en waarbij het bespreken van de implicaties qua omvang en inhoud vergelijkbaar is met de hiervoor besproken annotaties. De vraag rijst dan of het verschil in waardering te relateren is aan verschillen in de mate waarin en de wijze waarop de implicaties van de uitspraak worden besproken. Annotatie nr. 12 bespreekt hoe de door de rechter gewezen rechtsregel in de praktijk moet worden toegepast. In ongeveer anderhalve kolom (van de in totaal elf kolommen) bespreekt de annotator aan de hand van een rekenvoorbeeld hoe de regel toepassing vindt. Annotatie nr. 8 plaatst de besproken uitspraak, waarin de Hoge Raad een tussencategorie in een bepaald leerstuk introduceert, in een reeks van eerdere uitspraken. De annotator combineert de verschillende uitspraken en laat zien in welke situaties welke leer moet worden gehanteerd. Zo concludeert de annotator dat de introductie van een tussencategorie door de rechter niet tot gevolg heeft dat uitkomsten in eerdere uitspraken anders zouden zijn geweest. Ook merkt hij op dat in gevallen van complexe (meerpartijen)verhoudingen niet altijd duidelijk zal zijn welke criteria moeten worden toegepast.

Het beeld dat ontstaat, is dat het bespreken van gevolgen van de uitspraak voor de interpretatie of toepassing van het geldende (privaat)recht op zichzelf geen indicator is voor de kwaliteit van annotaties. Het kenmerk lijkt pas een indicator te zijn indien de passages waarin de implicaties worden besproken, van substantiële omvang blijken te zijn (niet enkele regels), waarbij de geschetste implicaties inhoudelijk¹⁶⁷ juist zijn en niet onmiddellijk voor de hand liggen.

Wijzen op belang van de uitspraak

In 3 van de 15 geanalyseerde privaatrechtelijke annotaties wijst de annotator op het belang van de uitspraak voor bijvoorbeeld de rechtspraktijk of het wettelijke

¹⁶⁷ Zie annotatie nr. 5, waarin het doortrekken van conclusies naar andere leerstukken mogelijk niet overtuigend was.

systeem (nr. 1, nr. 5, nr. 7). Aan alle drie de annotaties is door de beoordelaars (overwegend) een lage score toegekend. Mede gelet op de beperkte omvang van de passages waarin op het belang van de uitspraak wordt gewezen – het belang van de uitspraak wordt met een enkele zin aangegeven – en de aanwezigheid van andere kenmerken, is het onwaarschijnlijk dat de kwaliteit van annotaties positief dan wel negatief wordt beïnvloed door het louter aanstippen van het belang van de uitspraak.

Opvatting van andere auteur(s)toetsen / ter discussie stellen

De annotaties waarin het toetsen of ter discussie stellen van opvattingen van andere auteurs aan de orde is (nr. 4, nr. 8, nr. 10 en nr. 12), zijn alle (overwegend) positief beoordeeld door de beoordelaars. In annotatie nr. 4 wordt op twee punten in de annotatie een opvatting van een andere auteur ter discussie gesteld. De eerste keer licht de annotator in een halve alinea (drie relatief lange zinnen) toe waarom een redenering, die ook terugkomt in de wetsgeschiedenis, niet overtuigt. De tweede keer waarin de annotator een opvatting van een andere auteur ter discussie stelt, is dit sterk gerelateerd aan het gegeven dat de annotator zelf een criterium ontwikkelt en toetst welk criterium de rechter heeft toegepast. Dit maakt het moeilijk om voor deze annotatie het effect van het kenmerk op de kwaliteit van de annotatie vast te stellen, zeker nu uit de matrix volgt dat er nog andere factoren in het spel zijn.

In annotatie nr. 10 is de rol van het kenmerk beperkt. In deze annotatie constateert de annotator in één zin dat de gezichtspunten die hij uit een aantal uitspraken heeft achterhaald, in essentie overeenkomen met de gezichtspunten die in de literatuur zijn aangedragen. Niet zozeer deze vaststelling, maar veel meer het distilleren van gezichtspunten lijkt bepalend te zijn geweest voor de scores van de beoordelaars.¹⁶⁸

Dezelfde conclusie kan worden getrokken voor annotatie nr. 8 en nr. 12. Daar wordt, onder verwijzing naar de uitspraak, in een aantal regels een opvatting van een andere auteur bestreden. Het lijkt, gezien de marginale rol die het toetsen van andermans opvattingen hier heeft, onwaarschijnlijk dat dit wezenlijke invloed heeft gehad op de kwaliteitsoordelen.

Op basis van de geanalyseerde privaatrechtelijke annotaties kan, kortom, niet worden verondersteld dat het toetsen of ter discussie stellen van opvattingen van andere auteurs van invloed is op de kwaliteit van een annotatie. Voor zover die invloed er is, lijkt zij afhankelijk te zijn van de combinatie met andere factoren.

¹⁶⁸ Zie hiervoor onder 'Implicaties van de uitspraak'.

Onduidelijke passages constateren c.q. ophelderen/motiveringsgebrek opmerken

Het constateren en verduidelijken van onduidelijke passages of motiveringsgebreken is aan de orde in annotatie nr. 2 en nr. 8, welke annotaties beide als (overwegend) goed/uitstekend zijn beoordeeld. In annotatie nr. 8 merkt de annotator in één zin op dat het beter was geweest als de rechter een bepaalde keuze beter had gemotiveerd.

In annotatie nr. 2 komt het ophelderen van overwegingen twee keer en bovendien uitgebreider aan de orde. De eerste keer gaat het om het herformuleren van een moeilijk geformuleerde overweging, wat ongeveer een tiende van de annotatie beslaat (een derde pagina van in totaal drie en twee derde pagina's). De verduidelijking wordt gevolgd door een kritische bespreking en evaluatie van de betreffende overweging, wat een substantiëler deel van de annotatie beslaat. De indruk die gezien de omvang en de koppeling ervan aan het evaluatieve deel van de annotatie ontstaat, is dat het verduidelijken een meerwaarde kan hebben, maar vooral ook een opstap is naar een kritische bespreking van de uitspraak.

De tweede keer dat in annotatie nr. 2 sprake is van het ophelderen van een passage uit de uitspraak, constateert de annotator dat het door de rechter gebruikte criterium verwarrend is. Vervolgens doet de annotator een poging om een criterium te formuleren dat beter te hanteren is en aansluiting vindt bij de uitspraak. Hij komt hierbij tot de conclusie dat die oplossing niet de kritiek zal wegnemen die hij daarvoor in zijn noot beschrijft. Het ophelderen/verduidelijken van een onduidelijke passage hangt hier sterk samen met het bespreken van de implicaties van de uitspraak zoals de annotator die in een voorafgaande passage van de annotatie aan de orde heeft gesteld, en met het onderbouwd aangeven of de annotator het eens is met de uitspraak. Het is ook mogelijk dat het ophelderen opgaat in deze andere kenmerken.

Op grond van het vorenstaande zijn er geen sterke aanwijzingen gevonden op basis waarvan moet worden geconcludeerd dat het constateren van motiveringsgebreken en/of het ophelderen van onduidelijke passages een effect heeft op de kwaliteit van een (privaatrechtelijke) annotatie. Gezien de plaats, omvang en relatie met andere kenmerken, rijst het vermoeden dat het kenmerk geen wezenlijke bijdrage heeft geleverd aan de kwaliteit van de geanalyseerde annotaties, dat andere kenmerken daar meer invloed op hebben gehad, maar niet valt uit te sluiten dat het constateren of ophelderen van motiveringsgebreken of onduidelijkheden toch een rol speelt bij het bepalen van de kwaliteit van annotaties.

Gemotiveerd aangeven of de uitspraak in kwestie overtuigt

Het instemmen met, of bekritisieren van, een uitspraak is aan de orde in 6 van de 15 geanalyseerde annotaties, zowel bij annotaties die (overwegend) als goed/uitstekend zijn aangemerkt (nr. 2, nr. 12) als bij annotaties die dat niet zijn

(met name nr. 15) als bij annotaties die overwegend als matig/slecht zijn aangemerkt (nr. 5, nr. 6, nr. 14, nr. 15). Meestal (nr. 2, nr. 6, nr. 12, nr. 14) gaat het om opmerkingen ('dit is navolgenswaardig', 'hier valt veel voor te zeggen', 'dit oordeel is begrijpelijk') die terloops worden gemaakt bij het behandelen van de casus en de overwegingen. In annotaties nr. 5 en nr. 15 komt het kenmerk uitgebreider aan de orde, maar deze annotaties zijn overwegend als matig/slecht aangemerkt.

In annotatie nr. 2 gaat gemotiveerd aangeven of de uitspraak overtuigt bovendien op in het schetsen van implicaties. De annotator onderbouwt, met andere woorden, zijn stellingen ten aanzien van de uitspraak door de gevolgen ervan te bespreken, bijvoorbeeld door te laten zien tot welk resultaat de uitspraak leidt indien de besproken uitspraak wordt toegepast op nieuwe, al dan niet zelfbedachte gevallen.

Het voorgaande rechtvaardigt de conclusie dat het gemotiveerd aangeven of de uitspraak overtuigt op zichzelf geen (positief) effect heeft op de kwaliteit van annotaties. Alleen in combinatie met andere kenmerken lijkt er een effect te zijn, maar in deze gevallen is het effect vooral toe te rekenen aan die andere kenmerken.

Evaluatie van de uitspraak door deze af te zetten tegen andere denkbare uitkomsten

In een tweetal annotaties wordt een overweging uit de uitspraak afgezet tegen een ander mogelijk oordeel. Dit komt het meest uitgebreid aan de orde in annotatie nr. 2. In een passage, die ongeveer één alinea (zes relatief lange zinnen) beslaat, stelt de annotator onder verwijzing naar een eerdere uitspraak (zij het in een ander geval) dat als het criterium uit de uitspraak nader wordt aangescherpt, dit niet wegneemt dat er met de nodige soepelheid mee moet worden omgegaan. In annotatie nr. 12 bespreekt de annotator wat er zou zijn gebeurd als de rechter een bepaald argument had gehonoreerd. Dit onderdeel beslaat in beide annotaties ongeveer vier zinnen op de totale noot.

Hoewel het kenmerk in twee annotaties terugkomt die (overwegend) als goed/uitstekend zijn aangemerkt, kan niet worden vastgesteld dat het afzetten van de uitspraak tegen andere denkbare uitkomsten een indicatie is voor een goede of uitstekende noot. De omvang van de passages is beperkt. Bovendien hebben beide annotaties andere kenmerken waarvan de kans groter is dat deze de kwaliteit positief hebben beïnvloed. Dit geldt met name voor annotatie nr. 2, waar de indruk bestaat dat het schetsen van implicaties voor de toepassing of interpretatie van het recht een kernelement is van deze annotatie. Voor annotatie nr. 12 ligt het, gezien de eerdere bevinding, meer voor de hand dat de combinatie van het schetsen van de implicaties van de uitspraak en het afzetten van de

uitspraak tegen andere denkbare uitkomsten verklaart waarom deze annotatie (overwegend) goed is beoordeeld door de beoordelaars.

Overige kenmerken

Resteert een aantal kenmerken dat slechts in één van de privaatrechtelijke annotaties is aangetroffen. In annotatie nr. 4, die positief is beoordeeld door alle drie de beoordelaars, ontwikkelt de annotator een criterium en toetst hij vervolgens of de rechter dit criterium in de besproken uitspraak heeft toegepast. Dit toetsen vervult een centrale rol in de annotatie en beslaat dan ook bijna een vierde van de annotatie. De annotator komt tot het criterium door het af te zetten tegen een criterium zoals dat is voorgesteld door een andere auteur en door de consequenties van beide criteria te schetsen, mede in het licht van het beginsel dat in het betreffende rechtsgebied een belangrijke plaats inneemt. Gezien het evaluatieve karakter, de omvang van de passages waarin dit type evaluatie plaatsvindt en gezien de centrale plaats die het inneemt in de annotatie, is het mogelijk dat dit kenmerk een indicator is voor een kwalitatief goede of uitstekende noot. Omdat het kenmerk in slechts 1 van de 15 onderzochte annotaties terugkomt, moet echter de nodige voorzichtigheid worden betracht met deze conclusie.

Een ander kenmerk, dat ook in annotatie nr. 4 aan de orde is, is het kenmerk waarbij de annotator schetst welke (andere) mogelijkheden een van de procespartijen (in dit geval: contractspartijen) hadden om hun positie te verbeteren of te versterken. Het slot (deel van een alinea, vijf zinnen) van de annotatie gaat in op deze kwestie. Omdat de geschetste mogelijkheid los staat van de uitspraak, dat wil zeggen, ook had kunnen worden besproken indien de uitspraak niet zou zijn gedaan, is het de vraag in hoeverre dit kenmerk een (positief) effect heeft op de positieve waardering van annotatie nr. 4.

Het zoeken naar een verklaring waarom de rechter een bepaalde beslissing heeft genomen of deze heeft gemotiveerd, is aan de orde in annotatie nr. 8, die door twee van de drie beoordelaars als goed/uitstekend is beoordeeld. Dit kenmerk beslaat echter maar één zin op de totale annotatie (4,5 pagina's), zodat moet worden betwijfeld of dit kenmerk enig effect heeft gehad op de (positieve) beoordelingen van de beoordelaars, wetende dat het schetsen van implicaties ook aan de orde is in annotatie nr. 8, dat dit kenmerk meerdere alinea's beslaat en dat het schetsen van implicaties bij andere annotaties positief gerelateerd is aan de kwaliteit ervan.

In annotatie nr. 11 is voorts sprake van het toetsen van de beslissing aan het doel van de regeling of de regel in kwestie. Hierin besteedt de annotator één passage (drie zinnen) aan het toetsen van de uitspraak aan het doel van de besproken regel. Omdat de annotatie niet (overwegend) positief is beoordeeld door de beoordelaars, is er onvoldoende reden om aan te nemen dat dit kenmerk een indicator is voor de kwaliteit van annotaties. Dit geldt nog sterker voor het ken-

merk waarbij de annotator tracht te verklaren waarom de rechter heeft besloten zoals zij heeft besloten, zoals dat in annotatie nr. 8 aan de orde is. Het kenmerk betreft hier echter slechts één zin op de hele annotatie, zodat het vermoedelijk geen rol van betekenis speelt. In het licht van de bespreking van de hiervoor genoemde kenmerken hebben de implicaties die de annotator bespreekt waarschijnlijk een groter effect gehad op de (overwegend) positieve kwaliteitsoordelen van de beoordelaars.

Conclusie

De analyse van de privaatrechtelijke annotaties levert drie kenmerken op die als indicatoren kunnen dienen om de kwaliteit van privaatrechtelijke annotaties te bepalen. Een eerste indicatie is dat de annotatie niet verder gaat dan het weergeven van feiten, overwegingen, uitkomsten en/of het recht dat op het besproken geval van toepassing is. In die gevallen is de kans klein dat een annotatie positief gewaardeerd wordt door het juridisch-academisch forum.

Een tweede indicator is het bespreken van implicaties van de uitspraak. Dit kenmerk lijkt een positief effect te hebben op de kwaliteit van een privaatrechtelijke annotatie, mits de passages waarin de implicaties worden besproken, enige omvang hebben (bijvoorbeeld meerdere alinea's), inhoudelijk juist zijn en niet zonder meer voor de hand liggen. Het schetsen van implicaties kan op een heel praktisch niveau, onder meer door de gevolgen van de uitspraak voor de rechtspraak te bespreken ('hoe dient een advocaat te handelen?', 'hoe dient een rechter zijn uitspraak te motiveren?'), maar ook een meer theoretische bespreking komt voor, bijvoorbeeld door uit een hoeveelheid uitspraken (waaronder de besproken uitspraak) criteria of gezichtspunten te distilleren, of door aan de hand van nieuwe, fictieve gevallen na te gaan wat de gevolgen van de uitspraak zijn, dat als opstap kan dienen voor een bespreking van de wenselijkheid van de gevolgen, eventueel in het licht van alternatieve oplossingen of uitkomsten.

Een derde indicator is het ontwikkelen van een criterium en het toetsen of het criterium in de besproken uitspraak is toegepast. Hoewel dit kenmerk in één van de geanalyseerde, privaatrechtelijke annotaties is aangetroffen, heeft het gezien de centrale rol die het in die annotatie had, vermoedelijk een wezenlijk effect gehad op de positieve waardering van de drie beoordelaars.

Op basis van de geanalyseerde privaatrechtelijke annotaties kon daarentegen niet worden vastgesteld of het toetsen of ter discussie stellen van opvattingen van andere auteurs van invloed is op de kwaliteit van een annotatie, of dat het constateren van motiveringsgebreken en/of het ophelderen van onduidelijke passages een effect heeft. Voor zover die invloed er is, lijkt zij er alleen te zijn in combinatie met andere factoren.

Het vergelijken met andere, eerdere uitspraken, het wijzen op het belang van de uitspraak en het gemotiveerd aangeven of de uitspraak overtuigt blijkt bij privaatrechtelijke annotaties geen onderscheidend criterium te zijn aan de hand waarvan de kwaliteit kan worden bepaald, zeker niet indien de kenmerken niet of nauwelijks worden uitgewerkt.

Wat bij het doornemen van de annotaties, tot slot, opvalt, is dat de mate waarin één van de laatste twee indicatoren (bespreken implicaties en ontwikkelen/toetsen criterium), of een combinatie ervan, een substantieel deel van de annotatie vormt, dit de kans vergroot dat de annotatie positief wordt beoordeeld door deskundigen. Alleen annotatie nr. 5 voldoet niet aan dit beeld, maar dit kan worden verklaard doordat de annotator te veel heeft willen afleiden uit de uitspraak door de conclusies door te trekken naar andere leerstukken (nr. 5).

3 Analyse van strafrechtelijke annotaties

Ook voor de strafrechtelijke annotaties is een matrix opgesteld (Tabel 19). De patronen die hieruit volgen, worden hierna besproken, op een vergelijkbare wijze als bij de analyses van de bestuursrechtelijke en privaatrechtelijke annotaties. Alle zestien annotaties zijn geanalyseerd.

Tabel 19 – Matrix strafrechtelijke annotaties

Score (... van de ... beoordelaars merkt de annotatie als goed/uitstekend aan)													
22. Kritiek op voorstellen vanuit politiek													2 van de 2 (1.0)
21. Verwachting ontwikkeling													2 van de 2 (1.0)
20. Bespreking processtrategie													2 van de 2 (1.0)
19. Onduidelijkheid / motiveringsgebrek												1	2 van de 2 (1.0)
18. Annotator eens / oneens met uitspraak												1	2 van de 2 (1.0)
17. Annotator zoekt verklaring												0.5	2 van de 2 (1.0)
16. Argumenten voor en tegen beslissing												1	2 van de 2 (1.0)
15. Evaluatie adhv andere uitkomsten												0.5	2 van de 2 (1.0)
14. Evalueren beslissing adhv de praktijk													
13. Toetsen aan 'hoger' recht													
12. Toetsen aan doel van de regeling													
11. Toetsen aansluiting eerdere uitspraken												0.5	
10. Opvatting van andere auteur(s)													
9. Implicaties												0.5	
8. Criterium ontwikkelen en toetsen													
7. Nagaan welke leer is toegepast												1	
6. Opwerpen vragen													
5. Schetsen van kwesties													
4. Belang uitspraak												1	
3. Beschrijving geldend recht												1	
2. Beschrijving feiten/uitspraak												1	
1. Hoogste rechter												1	
ID													
1													1
2													2
3													2
4													2
5													2
6													2
7													2
8													2
9													0
10													1
11													1
12													1

13	1	1	1			1		1	1 van de 2 (0.5)
14	1	1	1		1				0 van de 2 (0.0)
15	1	1			1	0.5			0 van de 2 (0.0)
16	1	0.5	1	1		0.5			0 van de 2 (0.0)

Noot: Een score van '1' staat voor aanwezigheid van het betreffende kenmerk. Een score van '0.5' betekent dat het kenmerk is aangetroffen, maar dat de passage(s) van zeer geringe lengte zijn (één of enkele regels) in vergelijking met andere annotaties waarin het kenmerk is geobserveerd.

Toelichting bij categorieën (nummers corresponderen met nummers in tabel op kolomniveau)

1. Uitspraak gewezen door hoogste rechter
2. Beschrijving/samenvatting casus en/of uitspraak
3. Beschrijving geldend recht / verwijzen naar literatuur of jurisprudentie daarover
4. Wijzen op belang uitspraak voor rechtspraktijk of –theorie
5. Schetsen van kwesties die zich geregeld (in de praktijk) voordoen
6. Opwerpen vragen (zonder deze te beantwoorden)
7. Nagaan welke leer is toegepast (welke leer geldend recht is)
8. Criterium/criteria ontwikkelen en toetsen of de uitspraak aan dit criterium voldoet
9. Implicaties van de uitspraak bespreken (bv. voor toepassing/interpretatie geldend recht)
10. Opvatting van andere auteur(s) toetsen / ter discussie stellen
11. Toetsen of uitspraak aansluit bij eerdere uitspraken
12. Toetsen of beslissing van de rechter beantwoordt aan doel van de regeling
13. Toetsen van de uitspraak aan 'hogere' recht (beginselen / verdragsbepalingen / EHRM-rechtspraak)
14. Evalueren beslissing aan de hand van wat in de praktijk (van procesvoering) gebeurt / gangbaar is / gebruikelijk is
15. Evaluatie van de uitspraak door andere denkbare uitkomsten af te zetten tegen de gekozen uitspraak
16. Systematisch de argumenten voor en tegen de beslissing nalopen
17. Annotator geeft (onderbouwd) aan of hij/zij het eens dan wel oneens is met de uitspraak
18. Annotator zoekt naar verklaring voor uitkomst of gewezen rechtsregel
19. Onduidelijke passages constateren c.q. ophelderen / motiveringsgebrek opmerken
20. Aangeven of een procespartij beter een andere keuze had kunnen maken in het proces (bv. andere rechtsgrond invoeren) c.q. wat (andere) mogelijkheden zijn om rechtsposities te versterken / wat er zou zijn gebeurd als een procespartij een andere keuze zou hebben gemaakt
21. Verwachting voor ontwikkeling van leerstuk (volgt niet direct uit uitspraak, anders gecodeerd als 'schetsen implicaties')
22. Kritiek op voorstellen vanuit de politiek

Overzichten/samenvattingen/beschrijvingen

In alle van de strafrechtelijke annotaties die zijn geanalyseerd is sprake van een weergave van de casus, de uitspraak en/of het geldende recht dat van toepassing is op het besproken geval. Vier annotaties (nr. 9, nr. 14, nr. 15, nr. 16) zijn niet positief gewaardeerd. Kenmerkend voor deze annotaties – daarin wijken zij ook af van de annotaties die door beide beoordelaars als goed/uitstekend zijn aangemerkt – bestaan voor het overgrote deel uit bovengenoemde elementen. Weliswaar bevatten deze annotaties ook andere kenmerken – zo wordt het belang van de uitspraak omschreven (nr. 16), worden kwesties geschetst die zich geregeld voordoen (nr. 9), worden vragen opgeworpen (nr. 14, nr. 15), wordt vergeleken met eerdere uitspraken (nr. 9, nr. 16) en worden implicaties geschetst (nr. 15) – maar zoals hierna zal blijken, hebben deze kenmerken waarschijnlijk geen effect gehad op de negatieve scores die de beoordelaars aan de vier annotaties hebben toegekend. De kans is dan ook klein dat deze kenmerken de scores kunnen verklaren.

De lage waarderingen lijken vooral te moeten worden toegeschreven aan de omstandigheid dat de annotatie weinig andere informatie geeft aan de lezer dan een weergave van het arrest of van het geldende recht dat op het geval van toepassing is. Voorts lijken de beschrijvingen qua leesbaarheid en structuur niet wezenlijk af te wijken van die in annotaties die beter zijn beoordeeld. Er is dan ook geen reden om te veronderstellen dat de lage scores hieraan toe te schrijven zouden zijn.

Opvallend is dat annotaties nr. 10, nr. 11 en nr. 13 qua inhoud en opzet niet wezenlijk lijken te verschillen van de eerder genoemde annotaties. Ook deze annotaties bestaan voor het overgrote deel uit een weergave van de casus, het oordeel en/of het van toepassing zijnde geldende recht, een enkele keer aangevuld met een opmerking over het belang van de uitspraak (nr. 10) of een korte vergelijking met eerder verschenen uitspraken (nr. 10, nr. 13). Toch krijgen deze annotaties van één beoordelaar hogere scores dan de eerder genoemde annotaties. De meest voor de hand liggende verklaring is dat de annotator in de annotaties nr. 10, nr. 11 en nr. 13 onderbouwd aangeeft waarom hij/zij het al dan niet eens is met de uitspraak of onderdelen ervan. Dit kenmerk komt terug in de annotaties nr. 10, nr. 11 en nr. 13, maar niet bij de annotaties nr. 9, nr. 14, nr. 15 en nr. 16.

Omdat dezelfde beoordelaar de annotaties nr. 10, nr. 11 en nr. 13 als goed/uitstekend aanmerkt, rijst het vermoeden dat deze beoordelaar het al dan niet eens zijn met de uitspraak of onderdelen ervan als een relevant kenmerk van een goede/uitstekende annotatie beschouwt, terwijl dit bij de andere beoordelaar niet het geval is. De consequentie hiervan is dat de annotatie vermoedelijk door beide beoordelaars als matig/slecht zou zijn aangemerkt indien dit kenmerk afwezig zou zijn geweest.

Vergelijken met eerdere uitspraken

Het vergelijken met eerdere uitspraken is aan de orde in 12 van de 16 onderzochte annotaties. Dit kenmerk komt terug bij zowel annotaties die als goed/uitstekend zijn aangemerkt als bij annotaties die dat niet zijn. De vergelijkingen beperken zich doorgaans tot het aanstippen van overeenkomsten en verschillen ('deze houding bleek ook uit eerdere rechtspraak', 'overeind blijft ...', 'deze werkwijze kan bogen op een lange traditie. De rechter pleegt ...', 'in de eerdere zaak ging het niet alleen om ..., maar ook om ...'). Het vergelijken met eerdere jurisprudentie lijkt dan ook geen bruikbaar criterium te zijn om de kwaliteit van annotaties te beoordelen. Soms gaat de vergelijking verder dan een dergelijke, eenregelige signalering, maar van een echte vergelijking is dan geen sprake meer. Het vergelijken gaat dan over in het schetsen van implicaties ('is hiermee het ...-arrest achterhaald?', gevolgd door een analyse van de overeenkomsten en verschillen tussen de uitspraken). In één annotatie (nr. 13) wordt een uitgebreide vergelijking van meerdere alinea's gemaakt met een uitspraak uit een ander rechtstelsel, waarin zich een vergelijkbare zaak had voorgedaan. Het blijft hier overigens bij de constatering dat de Nederlandse uitspraak overeenkomt met die uit het buitenland, waarna deze buitenlandse uitspraak wordt beschreven.

Implicaties van de uitspraak

Het schetsen van de gevolgen van de uitspraak voor de toepassing of interpretatie van het geldende recht is aan de orde in 8 van de 16 strafrechtelijke annotaties die zijn geanalyseerd. Met uitzondering van annotatie nr. 15 zijn deze annotaties door beide beoordelaars als goed/uitstekend aangemerkt.

In annotatie nr. 15 bestaat het schetsen van de implicaties van de uitspraak uit één zin (op een noot van 2,5 pagina's). Daarbij komt dat de annotator zelf een slag om de arm houdt ('het zou wel eens zo kunnen zijn dat ...'). Dit doet vermoeden dat het schetsen van de implicaties in deze annotatie geen of nauwelijks een effect heeft gehad op de kwaliteit van deze annotatie.

Ook in de annotaties nr. 2, nr. 3 en nr. 4 zijn de passages waarin implicaties worden geschetst relatief kort (maar wel langer dan in annotatie nr. 15), en ligt een relatie tussen het schetsen van implicaties en de positieve beoordelingen niet onmiddellijk voor de hand. Annotatie nr. 2 bestaat voor twee derde (bijna twee van de ruim 2,5 pagina's) uit het systematisch nalopen en bespreken van de voor- en tegenargumenten voor een bepaalde uitkomst, wat overduidelijk de kern vormt van deze annotatie. Daarbij komt dat de implicaties die worden geschetst mede worden ontleend aan wat een andere auteur al had betoogd in een andere publicatie naar aanleiding van de uitspraak. Het is dan ook onwaarschijnlijk dat er in deze annotatie een relatie bestaat tussen het schetsen van de impli-

caties van de uitspraak en de positieve beoordelingen van de annotatie. Voor zover er een effect is, is het vermoedelijk klein.

Net als annotatie nr. 2, geldt voor annotaties nr. 3 en nr. 4 dat het schetsen van implicaties een bescheiden rol (één alinea) vervult, maar dat deze annotaties toch als goed/uitstekend zijn aangemerkt door beide beoordelaars. In annotatie nr. 4 schetst de annotator de implicaties van de uitspraak nadat hij is nagegaan welke van de twee benaderingen de rechter heeft toegepast in de uitspraak. Hij concludeert dat hoewel de rechter de ene benadering heeft toegepast, er toch nog ruimte blijft voor de andere benadering. De annotator vervolgt met een toelichting van hoe en waar die andere benadering een plaats heeft bij de beoordeling van gevallen. Hij schetst als het ware de gevolgen van de uitspraak en de keuze van de rechter voor een van de benaderingen.

In annotatie nr. 4 redeneert de annotator dat de omstandigheden van het geval beslissend zijn voor de beoordeling van gevallen zoals die aan de orde zijn in de uitspraak, dat deze omstandigheden niet nader worden gedefinieerd en dat daarom, bij het gebrek aan gezichtspunten, de vraag rijst of het leerstuk dat aan de orde is zelfstandige betekenis heeft. Het laatste komt neer op het schetsen van een implicatie van de uitspraak, namelijk dat de uitspraak laat zien dat de theorie niet deugt. Deze redenering lijkt de kern te vormen van de annotatie en lijkt te verklaren waarom de annotatie positief is beoordeeld.

Wel is duidelijk dat annotaties nr. 3 en nr. 4 de implicaties van de uitspraak uitgebreider beschrijven dan annotatie nr. 15, dat in de annotaties bovendien ook andere kenmerken aan de orde zijn, en dat deze omstandigheden samen het verschil in score tussen annotatie nr. 15 en annotaties nr. 3 en nr. 4 verklaren.

In de overige annotaties (nr. 5, nr. 6, nr. 7, nr. 8) beslaat het schetsen van implicaties van de uitspraak een substantieel deel van de annotatie, variërend van een vijfde (nr. 7) tot een derde (nr. 5, nr. 6, nr. 8) deel van de annotatie. In al deze annotaties is het schetsen van implicaties het dominante kenmerk, samen met het weergeven van de feiten, de uitspraak en/of het geldende recht dat op het geval van toepassing is. Niet uit te sluiten valt dat de andere, minder dominante kenmerken in deze annotaties ook een (positief) effect hebben op de kwaliteit van de annotaties, maar zoals hierna ook zal blijken, is dit effect vermoedelijk gering. Ten aanzien van het schetsen van implicaties kan dan ook worden vastgesteld dat dit, mits het enige omvang heeft, een positief effect heeft op de kwaliteit van annotaties.

Wijzen op het belang van de uitspraak

In 5 van de 16 geanalyseerde strafrechtelijke annotaties wijst de annotator op het belang van de uitspraak voor bijvoorbeeld de ontwikkeling van een leerstuk (nr. 1, nr. 16), voor het hanteren van een bepaalde regel door rechters en advocaten

(nr. 4), of als illustratie voor een goed functionerend strafrecht (nr. 10). Drie van deze annotaties zijn door beide beoordelaars als goed/uitstekend aangemerkt, één annotatie is door één beoordelaar als goed/uitstekend gekwalificeerd en één annotatie is door geen van beide beoordelaars als goed/uitstekend beoordeeld. Het is daarom, gezien de geringe omvang (meestal een enkele zin op de gehele annotatie) en gezien de aanwezigheid van andere, meer dominante kenmerken in annotaties die positief zijn beoordeeld, onwaarschijnlijk dat dit kenmerk een indicator is voor de kwaliteit van annotaties.

Nagaan welke leer is toegepast (welke leer geldend recht is) (evt. i.c.m. andere opvattingen beschrijven)

In drie annotaties, die alle door beide beoordelaars als goed/uitstekend zijn aangemerkt, beschrijft de annotator opvattingen van andere auteurs. In twee annotaties (nr. 1, nr. 4) gaat dit hand in hand met een analyse van welke leer is toegepast. In de annotaties nr. 1 en nr. 4 vormt dit de kern van de annotaties. In annotatie nr. 1 beslaat dit kenmerk anderhalve pagina van de in totaal drie pagina's, in annotatie nr. 4 ruim een halve pagina van de in totaal anderhalve pagina. In beide annotaties beschrijft de annotator twee dominante invalshoeken omtrent een bepaald strafrechtelijk leerstuk en gaat hij vervolgens aan de hand van de besproken uitspraak na bij welke invalshoek deze het beste aansluit. Zo laat de annotator in annotatie nr. 4 zien welke van de twee benaderingen is toegepast, en stelt en illustreert hij dat een bepaalde benadering een grotere verklaringskracht heeft dan een andere benadering die in de literatuur wordt aangehangen. Aan de hand van jurisprudentie en de besproken uitspraak laat de annotator zien waarom de betreffende benadering de uitspraken beter verklaart dan de andere benadering.

In annotatie nr. 5 dient de beschrijving van twee verschillende theorieën als opstap naar het doordenken van deze theorieën. Na de twee dominante stromingen te hebben geschetst, laat de annotator zien wat er zou zijn gebeurd als de rechter de andere stroming zou hebben gevolgd. Hij vervolgt met het schetsen van de implicaties van de keuze van de rechter om voor één van de benaderingen te kiezen. Dit alles neemt ongeveer ruim een derde van de annotatie (één pagina van de in totaal drie pagina's) in beslag.

Het toetsen welke leer is toegepast lijkt de kwaliteit van annotaties in positieve zin te beïnvloeden.

Kwesties schetsen die zich geregeld voordoen

In annotaties nr. 9 en nr. 12 beschrijft de annotator welke kwesties zich geregeld voordoen bij de toepassing van het leerstuk dat in uitspraak aan de orde is. In annotatie nr. 9 beslaat het schetsen van kwesties die zich geregeld voordoen een alinea op de hele annotatie (anderhalve pagina). De passage vormt de schakel

tussen een algemene beschrijving van het geval en de juridische vraag die in de uitspraak aan de orde is en nader wordt besproken in de annotatie. In annotatie nr. 12 begint de annotatie met de mededeling dat de vraag die in de uitspraak centraal staat, regelmatig aan de orde is. Op grond van de summiere omvang en gezien de relatief lage waarderingen voor de annotaties, ('0 uit 2' goed/uitstekend voor nr. 9, '1 uit 2' voor nr. 12), ligt het niet voor de hand dat het schetsen van kwesties die zich geregeld voordoen leidt tot een verhoging van de kwaliteit van (strafrechtelijke) annotaties. Er zijn voorts geen redenen om aan te nemen dat dit kenmerk een negatieve invloed heeft.

Opwerpen vragen (zonder deze te beantwoorden)

In annotatie nr. 14 en nr. 15 werpt de annotator aan het slot een (nr. 14) of enkele vragen (nr. 15) op die vervolgens niet verder worden uitwerkt. Het opwerpen van vragen beperkt zich hier dan ook tot één of enkele regels. Omdat er geen wezenlijke verschillen op dit punt zijn te constateren tussen deze twee annotaties, en gezien de lage waarderingen voor de annotaties nr. 14 en nr. 15, is het onwaarschijnlijk dat dit kenmerk een indicator is voor een kwalitatief goede of uitstekende annotatie.

Zoeken naar verklaring

Het valt op dat het zoeken naar verklaringen waarom een rechter een bepaalde uitspraak heeft gedaan of op welke gronden hij zijn of haar beslissing heeft gebaseerd, vooral aan de orde is in annotaties die door beide beoordelaars (nr. 3 en nr. 7) of door één van beide beoordelaars (nr. 10) als goed/uitstekend zijn aangemerkt. Dit doet vermoeden dat dit kenmerk een positief effect heeft op de kwaliteit van annotaties. Dit vermoeden blijkt echter moeilijk hard te maken omdat in diezelfde annotaties ook de implicaties van de uitspraak worden besproken. Hiervan werd eerder vastgesteld dat dit kenmerk waarschijnlijk een positief effect heeft op de kwaliteit van annotaties.

In annotatie nr. 10 is het schetsen van implicaties weliswaar niet aan de orde, maar het punt dat hier speelt, is dat de annotator in deze annotatie ook ingaat op de vraag of de uitspraak overtuigt. Hiervoor is vastgesteld dat de aanwezigheid van dit kenmerk voor één beoordelaar reden is geweest om een annotatie als goed/uitstekend aan te merken, zodat voor annotatie nr. 10 niet kan worden vastgesteld of de positieve beoordeling van deze annotatie het gevolg is geweest van het bespreken van waarom de uitspraak overtuigt of als gevolg van het verklaren waarom de rechter haar beslissing heeft genomen.

Het zoeken naar een verklaring voor de gewezen beslissing komt in annotatie nr. 3 niet uitgebreid aan de orde (ongeveer een alinea). In annotatie nr. 3 constateert de annotator dat de uitspraak een hoeveelheid vragen opwerpt. Hij bespreekt een aantal van deze vragen, gaat vervolgens op zoek naar een verkla-

ring voor open vragen en concludeert vervolgens dat de vragen het gevolg zijn van de casuïstische aanpak waar de rechter voor heeft gekozen. Deze passage wordt gevolgd door een passage waarin de annotator de implicaties schetst van deze casuïstische aanpak. Later gaat hij in een alinea in op een aantal onduidelijke passages van de uitspraak. Omdat geen van de kenmerken die de kwaliteit van annotatie nr. 3 zouden kunnen verklaren (zoeken naar een verklaring, schetsen van implicaties, onduidelijkheden constateren/ophelderen) dominant is in deze noot, is het niet mogelijk vast te stellen welke passages of welke kenmerken meer of minder van invloed zijn geweest op de positieve waarderungen. Mogelijk verklaren zij gezamenlijk positieve scores van de beoordelaars.

In annotaties nr. 7 en nr. 10 komt het kenmerk zijdelings aan de orde, dat wil zeggen in één (nr. 10) of enkele zinnen (nr. 7), zodat in deze gevallen moet worden betwijfeld dat het zoeken naar een verklaring voor de uitspraak een (wezenlijke) rol heeft gespeeld bij de kwaliteitsoordelen van de beoordelaars. Daarbij komt dat in annotatie nr. 7 het schetsen van de implicaties van de uitspraak een grotere en belangrijkere plaats in lijkt te nemen.

Concluderend kan worden gesteld dat er een zwakke indicatie is dat het zoeken naar verklaringen voor de uitspraak in positieve zin bijdraagt aan de kwaliteit van (strafrechtelijke) annotaties. Dit vermoeden volgt uit annotatie nr. 3, maar daar zijn ook andere factoren in het spel die eveneens de positieve waardering van de annotatie zouden kunnen verklaren, of beter: vermoedelijk dragen ook die andere kenmerken bij aan de positieve waarderungen.

Gemotiveerd aangeven of de uitspraak overtuigt

Het gemotiveerd aangeven of de uitspraak overtuigt, is aan de orde in 7 van de 16 bestudeerde strafrechtelijke annotaties. In twee van de annotaties beperkt dit zich tot een enkele regel of een korte passage, zoals in annotatie nr. 10 ('de uitspraak stemt tevreden') en nr. 11 ('de tekst van de uitspraak is een voorbeeld voor alle eerstejaars rechtenstudenten'). In de andere annotaties komt het gemotiveerd aangeven of de uitspraak overtuigt uitgebreider terug. Zo stelt de annotator in annotatie nr. 2 dat 'kanttekening moeten worden geplaatst bij dit uitgangspunt', waarna de annotator deze kanttekeningen bespreekt. Vergelijkbare passages zijn te vinden in annotatie nr. 6 en nr. 13. Annotatie nr. 12 gaat, weliswaar steeds kort, op ten minste tien plaatsen in de annotatie in op de beslissing van de rechter. Het patroon hierbij is dat de annotator begint met een korte weergave van de uitspraak, gevolgd door een kritische bespreking ('indien ... is deze regel redelijk', 'geen goede maatstaf', 'onbevredigend resultaat') zonder daarbij aan te geven (door te rekenen) waartoe de toepassing van de regel zou moeten leiden. Annotatie nr. 8, tot slot, is een vreemde eend in de bijt. De annotator geeft hier niet zozeer aan of hij het eens is met de uitspraak, maar met de conclusie

van de A-G, die zich beroept op eerdere publicaties maar deze volgens de annotator verkeerd heeft opgevat.

Annotaties nr. 2 en nr. 6 zijn door beide beoordelaars als goed/uitstekend aangemerkt. In deze annotaties neemt het gemotiveerd aangeven of de uitspraak overtuigt geen centrale plaats in.¹⁶⁹ Daarbij komt dat in de passages in deze annotaties waarin gemotiveerd wordt aangegeven of de uitspraak overtuigt vergelijkbaar zijn met de passages in de annotaties die minder goed zijn beoordeeld. Bovendien werd eerder al geconstateerd dat annotaties waarin gemotiveerd wordt aangegeven of de uitspraak overtuigt steeds als goed/uitstekend is aangemerkt door één van de twee beoordelaars. Dit geldt in ieder geval voor annotaties nr. 11, nr. 12 en nr. 13, en mogelijk ook voor annotatie nr. 10. Een andere verklaring voor het verschil in waardering tussen deze annotaties enerzijds en de annotaties nr. 9, nr. 14, nr. 15 en nr. 16 anderzijds kon niet worden gevonden, zo is eerder vastgesteld. Het ligt dan ook in de rede om te veronderstellen dat het gemotiveerd aangeven of de uitspraak overtuigt voor de ene beoordelaar wel, maar voor de andere beoordelaar geen reden is om annotaties met dit kenmerk als goed/uitstekend te kwalificeren.

Onduidelijke passages constateren c.q. ophelderen/motiveringsgebrek opmerken

Het constateren c.q. ophelderen van onduidelijke passages of van een motiveringsgebrek is aan de orde in annotaties nr. 1, nr. 2, nr. 3 en nr. 6. In annotatie nr. 2 is dit minimaal en blijft het bij signaleren ('vraag naar het waarom blijft helaas onbeantwoord'). Hier ligt het dan ook niet voor de hand dat dit kenmerk een effect heeft gehad op de kwaliteitsoordelen van de beoordelaars.

De overige annotaties hebben gemeenschappelijk dat dit kenmerk een alinea van de annotatie beslaat en dat de annotator aangeeft hoe de passage uit de uitspraak moet worden gelezen/geïnterpreteerd alsmede toelicht waarom dat op die manier dient te gebeuren. Zo lost de annotator in annotatie nr. 1 een innerlijke tegenstrijdigheid op en concludeert de annotator in annotatie nr. 3 dat sprake moet zijn van een vergissing in de formulering van de uitspraak. Omdat dit kenmerk niet terugkomt in de annotaties die niet als goed/uitstekend zijn aangemerkt, is het mogelijk dat het constateren c.q. ophelderen van onduidelijke passages en motiveringsgebreken een indicator is voor de kwaliteit van annotaties. Aan de andere kant geldt dat dit kenmerk vaak niet de kern lijkt te vormen van de annotaties; er zijn andere kenmerken in het spel die in alle van de genoemde annotaties een meer centrale plaats innemen. Of er een relatie bestaat tussen de kwaliteit van (strafrechtelijke) annotaties en het constateren c.q. op-

¹⁶⁹ In annotatie nr. 2 lijkt het verklaren waarom de rechter deze beslissing heeft genomen de kern van de uitspraak te vormen. In annotatie nr. 6 is het schetsen van implicaties de kern van de annotatie.

helderen van onduidelijke passages en motiveringsgebreken, is daarom moeilijk vast te stellen.

Overige kenmerken

Drie kenmerken komen terug in één van de annotaties. Een eerste kenmerk is dat waarin voor- en tegenargumenten op een systematische wijze worden behandeld, dat wil zeggen dat verschillende argumenten voor en tegen de beslissing worden nagelopen en worden besproken in termen van houdbaarheid. Dit kenmerk neemt in annotatie nr. 2 een centrale plaats in. Omdat deze noot door beide beoordelaars als goed/uitstekend is aangemerkt, gezien de centrale plaats van dit kenmerk – het beslaat twee van de drie pagina's van de annotatie – en gezien de beperkte omvang van de andere kenmerken die in de uitspraak aan de orde zijn die de positieve beoordelingen zouden kunnen verklaren, ligt het voor de hand dat de positieve scores van de beoordelaars (vooral) moeten worden toegeschreven aan de aanwezigheid van het systematisch nalopen en beoordelen van de voor- en tegenargumenten.

Een ander kenmerk dat in maar één van de bestudeerde (strafrechtelijke) annotaties is terug te vinden, is het uitspreken van een verwachting voor de ontwikkeling van een leerstuk (annotatie nr. 12). Dit heeft echter weinig om het lijf in deze annotatie. De verwachting bestaat eruit dat de regels nog wel ooit zullen worden aangepast en worden verfijnd omdat men strenger wil zijn of omdat de regels anderszins dienen te worden heroverwogen. Het is dan ook onwaarschijnlijk dat er een relatie bestaat tussen dit kenmerk en de kwaliteit van annotaties.

Het derde kenmerk dat in alleen in annotatie nr. 5 is aangetroffen, is het afzetten van de uitkomst tegen alternatieve uitkomsten ('wat als de rechter anders had beslist?'). Deze bespreking betreft één alinea en dient als opstap naar het bespreken van de implicaties van de uitspraak. Dat laatste, het bespreken van de implicaties, vormt de kern van de annotatie, zodat niet is vast te stellen of het afzetten van de uitspraak tegen alternatieve uitkomsten, een effect heeft op de kwaliteit van annotaties. Dit is ook niet uit te sluiten, maar op basis van deze gegevens kan daar geen uitspraak over worden gedaan.

Conclusie

Van een aantal kenmerken kan worden vastgesteld, in ieder geval worden vermoed, dat het kan worden gerelateerd aan de kwaliteitsoordelen van de onderzochte (strafrechtelijke) annotaties. Een eerste vaststelling is dat indien een annotatie weinig andere informatie geeft aan de lezer dan wat ook uit de uitspraak of uit handboeken te halen valt, de kans klein is dat een annotatie als goed/uitstekend wordt aangemerkt.

Voor het schetsen van implicaties van de uitspraak voor de toepassing of interpretatie van het recht geldt dat dit vermoedelijk een positief effect heeft op

de kwaliteit van annotaties. Een voorwaarde is wel dat de passage(s) waarin de implicaties worden geschetst, een behoorlijke omvang heeft (hebben) (ten minste meer dan een enkele alinea). Ook bij een beperkte omvang kan een annotatie positief worden gewaardeerd, maar meestal zijn er dan tevens andere kenmerken in het spel die de kwaliteit van de annotatie vergroten.

Het effect van het nagaan van welke leer is toegepast (in combinatie met opvattingen van anderen beschrijven) op de kwaliteit van (strafrechtelijke) annotaties is moeilijk vast te stellen. Het nagaan welke leer is toegepast gaat vaak hand in hand met andere kenmerken, zoals het beschrijven van opvattingen van anderen, het schetsen van implicaties of het schetsen van alternatieve uitkomsten. Toch lijkt de aanwezigheid van dit kenmerk (mits van een behoorlijke omvang) een indicatie te zijn van een kwalitatief goede noot. Het kenmerk vormt de kern van een aantal annotaties en komt uitsluitend terug in annotaties die door beide beoordelaars als goed/uitstekend zijn aangemerkt.

In een aantal van de onderzochte annotaties was sprake van het zoeken naar een verklaring voor de uitkomst of de gewezen rechtsregel. De (positieve) relatie met de kwaliteit van annotaties is echter moeilijk vast te stellen. In de annotatie waaruit dit vermoeden is afgeleid, waren ook andere factoren in het spel die vermoedelijk ook hebben bijgedragen aan de positieve waardering van de betreffende annotatie.

Een kenmerk dat in slechts één van de onderzochte annotaties terugkomt, maar waarvan het effect voor de hand ligt gezien de plaats en functie ervan binnen de annotatie, is het op systematische wijze bespreken van voor- en tegenargumenten, waarbij het niet alleen gaat om het bespreken van de argumenten, maar ook op het beoordelen van de houdbaarheid van de verschillende argumenten.

Van een tweetal kenmerken kon niet overtuigend worden vastgesteld of er een (positieve) relatie bestaat met de kwaliteit van annotaties. Het gaat om het kenmerk waarbij de uitkomst wordt vergeleken met de uitkomst indien de rechter anders zou hebben besloten, en om het kenmerk waarbij onduidelijke passages en/of motiveringsgebreken worden geconstateerd c.q. opgehelderd. In de annotaties waarin dit aan de orde was, waren ook andere kenmerken in het spel, die vaak bovendien een meer centrale rol leken te hebben dan de hiervoor genoemde kenmerken.

Verder is gebleken dat een aantal kenmerken waarschijnlijk niet te relateren is aan de kwaliteit van een (strafrechtelijke) annotatie. Het gaat om het vergelijken met eerdere uitspraken, het wijzen op het belang van de uitspraak, het schetsen van kwesties die zich geregeld voordoen in de rechtspraktijk of in jurisprudentie, het opwerpen van vragen (zonder deze te beantwoorden) en om het uitspreken van een verwachting voor de ontwikkeling van een leerstuk.

Voor één kenmerk geldt, tot slot, dat dit voor de ene beoordelaar wel, maar voor de andere beoordelaar geen criterium is geweest om annotaties als goed/uitstekend aan te merken, namelijk het gemotiveerd aangeven of de uitspraak overtuigt.